Spoken word production in Norwegian-English bilinguals

Malin Toften Mangersnes

UiA University of Agder

Doctoral Dissertations at the University of Agder 443

Spoken word production in Norwegian-English bilinguals

Mangersnes, Malin Toften

Spoken word production in Norwegian-English bilinguals

Investigating effects of bilingual profile and articulatory divergence

Dissertation for the degree philosophiae doctor (PhD)

University of Agder
Faculty of Humanities and Education
2023

Doctoral dissertations at the University of Agder 443

ISSN:1504-9272
ISBN: 978-82-8427-160-6

Printed by Make! Graphics
Kristiansand, Norway

Acknowledgements

First, I would like to thank my supervisors **Allison Wetterlin** and **Linda Wheeldon**. Your knowledge and guidance have been instrumental to the development and execution of this project, and a source of motivation throughout. The thesis is now (finally!) finished, much due to your continued support and contagious enthusiasm. I truly appreciate our meetings and discussions, your kindness, and the environment you both have created in the lab and beyond.

I would also like to thank the Faculty of Humanities and Education and the Department of Foreign Languages and Translation at the University of Agder for making this project possible.

A big thank you goes to the participants for spending hours in the Experimental Linguistics Lab. Without your data there would be no thesis, and I am very grateful for your contributions.

Luckily for me, the lab is filled with friendly, bright, and helpful people. I would particularly like to thank: **Jan Zandhuis** for general technical support and developing the forced aligner, **Lone Sunnset** for questionnaire data and phonological neighbourhood densities, **Maud Pelissier** for helpful advice - especially regarding statistics, and **Beatrice**, **Bjørn**, **Dag**, **Mikael**, **Sophie**, and **Susanne** for contributing to the experiments. Finally, a collective thank you to everyone who has passed through the lab, whether to work, study, participate in piloting, keep the lab in shape, or simply provide a welcome distraction, I appreciate you all.

I am thankful for the people I have met both at the University of Agder and at other universities. Thank you for the productive discussions and the useful input, and especially **Mila Vulchanova** for thorough comments on an earlier version of this thesis. To my colleagues, students, and friends at the university, thank you for making everyday life more delightful and interesting. To the people at Knausen, in the coffee corner, the E-corridor, and in ML vei, thank you for the inspiring talks, laughs, silliness, and great memories. Thank you, **Verena**, for welcoming me to the university and being a brilliant doctoral accomplice.

Thank you, **Astrid Aksnes**, for proofreading this beast, and thank you **Karin**, **Marit**, and **Tina** for proofreading in Norwegian and helpful suggestions.

Finally, I want to thank my family and friends for their encouragement, visits, and being lovely people in general. To my family, I am immensely grateful for your support in the final stages of writing this thesis, thank you for making it easier. You are the best!

Malin Toften Mangersnes
Trondheim, June 2023

Summary in Norwegian

Avhandlingen undersøker hvordan individuelle forskjeller i tospråklig profil (språklig bakgrunn, språkferdigheter og språkeksponering) og artikulasjon påvirker tospråklig ordproduksjon. Når tospråklige produserer ord er begge språkene de snakker aktive til en viss grad, noe som tilsier at mekanismer for språkkontroll er nødvendige for å sikre at det tiltenkte ordet blir produsert. Disse kontrollmekanismene har blitt undersøkt i eksperimenter hvor deltakerne veksler mellom språkene. Tidligere studier har vist at generelle språkferdigheter påvirker hvor raskt og nøyaktig tospråklige produserer ord (f.eks., Costa & Santesteban, 2004). Tospråklig ordproduksjon påvirkes også, stort sett i positiv forstand, av at ord har lik form og betydning i første-(S1) og andrespråket (S2) (f.eks., Costa et al., 2000; Hoshino & Kroll, 2008). Siden effekten av ord som er like i form og betydning på tvers av språk (heretter *kognater*) delvis skyldes fonologisk likhet mellom språkene, og graden av likhet kan variere, fokuserer denne avhandlingen spesielt på individuelle forskjeller relatert til fonologi og uttale. I denne avhandlingen ble tospråklig ordproduksjon undersøkt i to eksperimenter hvor deltakerne fikk se bilder av ulike objekter og ble bedt om å navngi objektene så raskt som mulig. Først navnga deltakerne bilder i en ettspråklig kontekst, hvor bilder i S1 og S2 ble navngitt hver for seg, og deretter i en vekslingskontekst hvor de vekslet mellom språkene. Individuelle forskjeller ble undersøkt ved hjelp av flere metoder.

Målingene av individuelle forskjeller i tospråklig profil og språklige ferdigheter ble analysert og sammenlignet med hverandre i kapittel 4. Et spørreskjema for tospråklige ble utvidet med spørsmål om uttale, fonologi og veksling mellom språk. Fire tester på norsk (S1) og engelsk (S2) målte språkferdigheter (vokabular og staving) og fonologisk prosessering (elisjon og *serial nonword recognition* - SNWR). I en faktoranalyse av svarene på spørreskjemaet ble variabler forbundet med fonologi og uttale (spesielt i S2) plassert i samme faktor, noe som tyder på at de nye spørsmålene bidrar til beskrivelsen av tospråklige profiler. Samtidig var det noe sammenfall med faktorer som inneholdt variabler forbundet med generelle språkferdigheter og språkeksponering. Faktorene og selvrapporterte språkferdigheter ble så sammenlignet med testresultatene. Den ene fonologiske testen, elisjon, korrelerte med deltakernes vurderinger av egen uttale i både S1 og S2. Faktoren som var forbundet med S2-uttale korrelerte med S1-elisjon, men ikke med de andre fonologiske testene.

Effekten av individuelle forskjeller i tospråklig profil på tospråklig ordproduksjon ble undersøkt i kapittel 5. Reaksjonstider og feilfrekvens ble målt i to eksperimenter hvor deltakerne navnga bilder i en ettspråklig kontekst og en språkvekslingskontekst. Individuelle forskjeller ble målt med to faktorer fra analysen i kapittel 4, den ene forbundet med S2-uttale (*L2 accent and interest*) og den andre med generelle S2-ferdigheter (*General L2 proficiency*), og deltakernes resultater fra en flanker-test. I begge eksperimentene var halvparten av ordene kognater. Tidligere rapporterte effekter av ordtype (kognater eller ikke-kognater), svarspråk (S1 eller S2) og språkveksling ble reprodusert med en ny gruppe tospråklige deltakere som behersker begge språkene godt. I eksperimentet med språkveksling ble det også reprodusert symmetriske vekslingskostnader (*symmetric switch costs*), det vil si at det var like krevende å bytte fra S1 til S2 som S2 til S1. Resultatene fra eksperimentet med språkveksling støtter tidligere funn av global S1-inhibisjon i språkkontroll. I den ettspråklige konteksten var reaksjonstidene raskere i S1, mens reaksjonstidene var raskere i S2 i eksperimentet med språkveksling (*reversed dominance*). I begge eksperimentene var det en effekt av ikke-lingvistisk oppmerksomhetskontroll (målt med flanker-testen). Den positive effekten av bedre oppmerksomhetskontroll var større for ikke-kognater enn kognater i eksperimentet med ettspråklig kontekst, mens det var en generell positiv effekt i eksperimentet med språkveksling. Det var ingen signifikante effekter av individuelle forskjeller i generelle S2-ferdigheter. Høyere verdier av faktoren forbundet med S2-uttale var assosiert med raskere reaksjonstider generelt i eksperimentet med ettspråklig kontekst, men en interaksjon mellom faktoren, ordtype og svarspråk indikerte at dette ikke var tilfellet for ikke-kognater i S2. I eksperimentet med språkveksling var det marginale effekter av faktorene. Disse funnene bekrefter den positive effekten av kognater i ordproduksjon, men tyder ikke på at effekten påvirkes av individuelle forskjeller slik de er målt i denne avhandlingen.

I kapittel 6 undersøkte jeg sammenhengen mellom individuelle forskjeller i artikulasjon, tospråklig profil, språkferdigheter og ordproduksjon. To ulike mål på forskjeller i artikulasjon ble regnet ut for to par like, men ikke identiske, S1- og S2-vokaler, /uː/-typen (S1 /ʉː/ og S2 /uː/) og /ʌ/-typen (S1 /œ/ og S2 /ʌ/). Generelt var det mange sammenligninger og få signifikante resultater. De ulike målene gav også ofte forskjellige resultater. Ingen av dem var relatert til selvrapporterte uttaleferdigheter. Forskjeller i produksjonen av /uː/-typen vokaler var relatert til S1-SNWR, men det var ingen andre sammenhenger mellom forskjeller i artikulasjon og de fonologiske testene. Noen av faktorene forbundet med tospråklig profil var relatert til individuelle forskjeller i artikulasjonen av /ʌ/-typen vokaler. I eksperimentene med ordproduksjon var det kun begrensede effekter av individuelle forskjeller i artikulasjon. Effektene som ble observert i den ettspråklige konteksten var i retning av raskere produksjon med større akustiske forskjeller mellom S1- og S2-vokaler, bortsett fra for ikke-kognater i S1.

Dette prosjektet legger grunnlaget for videre undersøkelser av hvordan individuelle forskjeller i tospråklig profil, spesielt når det gjelder fonologi og uttale, og artikulasjon påvirker tospråklig ordproduksjon. Resultatene tyder på at individuelle forskjeller relatert til fonologi og uttale i noen grad påvirker ordproduksjon, men for tydeligere svar må metodene videreutvikles.

Summary in English

This thesis investigates the effects of individual differences in bilingual profile and articulatory divergence on bilingual spoken word production. During bilingual word production both languages are to some degree active, and control mechanisms are therefore required to produce the intended word. These control mechanisms have been investigated in language switching experiments, and general language proficiency differences have been shown to modulate naming behaviour (e.g., Costa & Santesteban, 2004). The motivation for additionally focusing on individual differences in the domains of phonology and accent in the current study came from observations of cognate effects in word production (e.g., Costa et al., 2000; Hoshino & Kroll, 2008). Cognates are words that share form and meaning across languages and often facilitate word production. Cognate effects have been partly attributed to cross-linguistic phonological similarity, but the degree of similarity between cognates is subject to individual variation. In this thesis the classic psycholinguistic paradigms of picture naming and language switching were used to assess bilingual word production. A variety of measures were employed to assess individual differences in bilingual profile, proficiency, and articulation.

In the first of three experimental chapters (Chapter 4) the different measures of bilingual profile and proficiency employed in this thesis were analysed and compared. For this thesis a bilingual profile questionnaire was augmented to obtain more information relating to accent, phonology, and language switching. Paired language tests in Norwegian (L1) and English (L2) served as objective measures of proficiency (vocabulary and spelling) and phonological processing (elision and serial nonword recognition). The questionnaire was factor analysed and questions relating to phonology and accent grouped together, suggesting that the augmentations contribute to the assessment of bilingual language profile. However, there was also some overlap with factors relating to general language proficiency and language exposure. Both the factors and self-ratings of proficiency were compared to language task performance. Of the phonological tests, elision scores, but not serial nonword recognition (SNWR) scores, correlated with self-ratings of pronunciation proficiency in both languages. The *L2 accent and interest* factor correlated with L1 elision, but none of the other phonological measures.

In Chapter 5 the effect of individual differences in bilingual profile and proficiency on word production and language control was investigated. Two of the factors extracted from the questionnaire, *General L2 proficiency* and *L2 accent and interest,* and performance on a flanker task were added as predictors of latencies and accuracy in L1 and L2 picture naming and a language switching task. Both tasks included a cognate status manipulation. Previously reported main effects of cognate facilitation, naming language, and trial type on naming latencies were replicated with a new group of bilinguals, as well as findings of symmetrical switch costs for relatively proficient bilinguals. The results also support and extend findings of global L1 inhibition in language control. A reversed dominance effect was observed in the language switching task, where naming was faster in the L2 than in the L1. Individual differences in the non-linguistic measure of attentional control interacted with cognate status in the picture naming task, where the effect of faster production with better attentional control was larger for noncognates than cognates, while there was a general benefit in the switching task. There were no significant effects of the *General L2 proficiency* factor. In the picture naming task, higher values of the *L2 accent and interest* factor were associated with faster naming in general. However, an interaction with cognate status and naming language indicated that this was not the case for L2 noncognates. In the switching task both predictors were involved in marginal interactions. These findings confirm the facilitatory effects of cognates in naming. However, they do not appear to be strongly modulated by the measures of individual differences assessed here.

The last experimental chapter (Chapter 6) investigated how individual differences in articulation related to bilingual profile, proficiency, and word production. Two measures of articulatory divergence were calculated for two pairs of similar, but not identical, L1 and L2 vowels, the /uː/-type vowels (L1 /ʉː/ and L2 /uː/) and the /ʌ/-type vowels (L1 /œ/ and L2 /ʌ/). Overall, there were many comparisons and few significant results. One clear observation was that different effects were found depending on which divergence measure was used. The divergence measures did not relate to self-ratings of pronunciation proficiency but there were some significant relationships with test scores. Higher L1 SNWR scores were associated with more divergent productions of the /uː/-type vowels, but there were no other links between the divergence measures and the phonological tests. The /ʌ/-type vowels, but not the /uː/-type vowels, were related to aspects of bilingual profile. Finally, there were limited effects of articulatory divergence on bilingual language production. The observed effects in the picture naming task were in the direction of faster naming with more divergent production, apart from for L1 noncognates.

This project constitutes a first step in the investigation of individual differences in bilingual profile, particularly relating to the domains of phonology and accent, and articulatory divergence on bilingual word production. The results demonstrate a role for individual differences in the domains of phonology and accent, but for clearer answers the measures used need to be further developed.

Contents

List of Figures

List of Tables

1 Introduction

A large proportion of language users are bilingual, and some researchers have claimed that over 50% of the world's population is bilingual or multilingual (Marian & Shook, 2012). Based on self-reported data 54% of EU citizens, and 74% of 15 – 24 year olds, are able to hold a conversation in at least one language in addition to their native language (European Commission, 2012). Languages can be learned throughout life, in different learning contexts and with different outcomes. This means that there is a great deal of variation in bilingual language proficiency and language experience which has consequences for both first and second language representation and processing.

Evidence suggests that both languages spoken by a bilingual are activated during language processing, i.e. non-selective activation (e.g., Gullifer et al., 2013; Poarch & van Hell, 2012; Y. J. Wu et al., 2013), yet bilinguals are generally successful in speaking the target language while supressing the other(s) (Gollan et al., 2011). This means that a control mechanism must be active during bilingual speech production (e.g., Declerck & Philipp, 2015). Language control can be studied using language switching tasks. An asymmetric switch cost (longer latencies associated with switching from a weaker language to a stronger language than vice versa) has been observed for bilinguals who are less proficient in one of their languages (e.g., Meuter & Allport, 1999), while this difference is attenuated in bilinguals who are highly proficient in both languages (e.g., Costa & Santesteban, 2004).

The representation of sound structure is potentially an important factor for investigating language non-selectivity as there is evidence that cross-linguistic activation extends to phonological form (e.g., Christoffels et al., 2007). In picture naming, words sharing form and meaning across languages (i.e., cognates[1], for example English *bus* (/bʌs/) and Norwegian *buss* (/bʉs/)) are produced faster than translation equivalents with different forms in each language (i.e., noncognates, for example English *chair* (/tʃɛː/) and Norwegian *stol* (/stuːl/)). Differences in

[1] Cognates, when mentioned in this thesis and in psycholinguistic research, refer to words that share form and meaning cross-linguistically. In philology this definition is reserved for words sharing the same origin, thus excluding loanwords. This is discussed briefly in Section 5.2.1.

language proficiency have also been shown to modulate this effect, such that cognate benefits are larger for a weaker language compared to a stronger language (e.g., Costa et al., 2000; Rosselli et al., 2014).

One aspect of language proficiency that has received sparse attention in this context is phonology and phonetic production. In language acquisition research, the learnability and representation of second language (L2) phonology, especially in terms of age-related constraints, has received a lot of attention (e.g., Long, 1990; Yeni-Komshian et al., 2000). However, the nature of L2 phonological representations and the consequences for language processing remain understudied. In this project, measures of language proficiency and acoustic measures of articulation are combined with classic paradigms from psycholinguistics to investigate the relationship between these individual difference measures, word production and language control.

As mentioned above, cognates can facilitate bilingual word production and cognate effects are modulated by proficiency differences (or lack thereof) between the L1 and L2. Cognates differ in their degree of form overlap between languages (Schepens et al., 2013) and there is some evidence that cognate effects are modulated by the degree of form overlap. For instance, in visual lexical decision cognate effects are stronger when there is more orthographic overlap (Dijkstra et al., 2010) and a priming study using different scripts (Japanese and English) found that cognate effects were stronger when the phonological similarity was higher (Nakayama et al., 2014). Importantly, there is also variation in cognate similarity between individuals speaking the same languages, both due to differences in L1 dialects and in L2 sound representations.

One central model of L2 sound learning and representation is the Speech Learning Model (Flege, 1995, 2007). Within this framework L1 and L2 representations are part of two separate subsystems existing in a common phonological space, and there is a strong relationship between perception and production. New speech sound categories can be established for an L2, depending on the perceived degree of difference from existing L1 sound categories. When L2 representations are not formed, L2 sounds may continue to be processed though the L1 system or a merged category may be formed that is dissimilar to monolingual representations in either language. In this thesis the acoustic difference between L1 and L2 vowels serve as a measure of articulatory divergence.

Defining and measuring competence in the domains of phonology and accent is complex. Self-ratings in other domains are frequently used to assess proficiency and tend to correlate with language performance (e.g., Marian et al., 2007). Direct self-ratings of phonological proficiency are more difficult to obtain as phonological representations are abstract, and are likely affected by individual differences, for instance in phonological memory (e.g., Aliaga-Garcia et al., 2010) and phonological awareness (e.g., Saiegh-Haddad, 2019). Questionnaires frequently contain a question asking participants to rate their L2 accent. Several factors may lead speech to be perceived as accented, but in the current study segmental differences are of particular interest. In the current study, self-ratings of pronunciation proficiency and interest and attitudes to accent are collected, in addition to performance on phonological tests and measures of articulatory divergence. This information will be used to quantify individual differences in what will be referred to as *the domains of phonology and accent*.

In this thesis, I examined data collected from 60 native speakers of Norwegian with English as a second language. The same participants completed a set of experiments in both their L1 and L2 in two different experimental sessions, and in the final session they completed a language switching task. They had similar educational backgrounds and a similar age of language acquisition. The aim is to investigate the effects of individual differences in bilingual profile and articulatory divergence on bilingual spoken word production.

1.1 Overview of thesis structure

The thesis is divided into three experimental chapters (Chapters 4 - 6), which are preceded by a theoretical introduction (Chapter 2) and an overview of general methodology and the motivation for each study component (Chapter 3). The findings from all experiments are summarised and discussed in Chapter 7.

Chapter 2 introduces the key concepts, theories and models that form the background to this research. The term bilingual is briefly discussed and models of both monolingual and bilingual word production are introduced. Models of bilingual language control and studies of bilingual language switching are briefly reviewed, before phonological representations and aspects of L2 perception and production are discussed. The theoretical introductions to the experimental chapters (4 – 6) discuss relevant key aspects of these theories and empirical findings in more detail.

Chapter 3 describes the motivation for the current project and the key issues addressed in this thesis. It also provides an overview of the experimental components and the general methodology. This study is designed to investigate individual differences in bilingual language production using several types of data including a bilingual profile questionnaire, objective tests of language proficiency, psycholinguistic experiments with spoken word production and language switching, and acoustic measures of L1 and L2 articulatory divergence.

Chapter 4 investigates the relationships between self-reported proficiency, bilingual profile measures and tested proficiency, with a particular interest in the domains of phonology and accent. Sixty participants in this experiment completed an extensive questionnaire, augmented with questions relating to L2 accent and phonological awareness, and four language tests, both in their L1 and L2. To determine the relationship between variables in the domains of phonology and accent and other aspects of bilingual language profile, and to reduce the dataset, the questionnaire data were factor analysed. Variables relating to accent and phonology generally grouped together. The relationships between the factors, objective tests, and proficiency ratings were also examined. Generally, language tests were found to be more strongly related to L2 self-reported proficiency and L2-related factors than to L1. Self-rated pronunciation proficiency correlated with phoneme elision scores in both languages, but not serial nonword recognition, and L1 elision scores correlated with the factor comprised of accent and phonology related variables.

Chapter 5 reports on two experiments designed to investigate the role of individual differences in bilingual spoken word production and language control. Cognate status was manipulated in a picture naming task and a language switching task. Main effects of cognate facilitation and language are replicated with a new bilingual population in an experiment with more stimuli and fewer repetitions of each picture than is typically used in these studies. Both tasks show robust main effects of cognate facilitation. In simple picture naming, L1 responses were faster compared to L2, while L1 responses were slower in the switching task. The data were collected in three separate sessions. In the two first sessions participants completed the simple picture naming task in English and Norwegian on separate days (language order was counter-balanced), before they completed the switching task in the final session. Significant effects of language order in the simple picture naming sessions (i.e., Norwegian or English first) were observed in both the picture naming task and the switching task. To investigate the effects of individual differences, performance on a flanker task, and the *General L2 proficiency* and *L2 accent and interest* factors obtained from the questionnaire were added as predictors to models of the naming and switching data. Better attentional control was associated with faster naming in both languages and in both tasks, but in the simple picture naming task the benefit was larger for noncognates. There were no significant effects of *General L2 proficiency* but higher values of the *L2 accent and interest* factor scores were associated with faster naming in the picture naming task. In the switching task, both predictors were involved in marginal interactions, but there were no significant effects.

In Chapter 6, the focus turns to assessing the relationship between individual differences in articulation, proficiency, bilingual profile, and bilingual word production. Individual differences in articulation are obtained in an acoustic analysis of articulatory divergence between two pairs of similar, but not identical, L2 and L1 vowels (L2 /uː/ - L1 /ʉː/ and L2 /ʌ/ - L1 /œ/). The divergence measures did not relate to self-ratings of pronunciation proficiency. Higher scores on the L1 serial nonword recognition task were associated with more divergent productions of the /uː/-type vowels, but there were no other links between the divergence measures and phonological tests. The /ʌ/-type vowels, but not the /uː/-type vowels, were related to aspects of bilingual profile. Finally, there were limited effects of articulatory divergence on bilingual language production. The observed effects were in the direction of faster naming with more divergent production, apart from for L1 noncognates in the picture naming task. One clear observation is that different effects were found depending on which divergence measure (Pillai score or BA) and which vowel pair were used to quantify individual differences in articulation.

Chapter 7 is the final chapter of this thesis. Herein, key findings are summarised and discussed. Overall, the results suggest a limited role for individual differences

in bilingual profile, assessed by two factors from the questionnaire, and articulatory divergence in bilingual word production. Significant effects were observed in simple picture naming, but not naming with language switching. Faster naming was generally associated with higher values of the *L2 accent and interest* factor and more divergent articulation. The complex relationships between bilingual profile factors, objective language tests, acoustic divergence measures and self-rated proficiency are discussed, along with future suggestions.

2 Theoretical overview

2.1 Introduction

In this chapter I introduce the key concepts, theories and models that provide the foundation for the current investigation into the effects of individual differences in bilingual profile and articulatory divergence on bilingual spoken word production. I start with a short discussion of *bilingualism* and how this term is understood and used in this thesis, then models of monolingual and bilingual word production are introduced. The next section discusses some key findings on bilingual language switching and models of bilingual language control. Finally, phonological representations in L1 and L2, as well as the link between and L2 perception and production are briefly discussed.

2.2 Defining bilingualism

When looking at studies of bilingual language processing it quickly becomes apparent that the term *bilingual* is used to describe many different types of language learners and users. C. Baker (2001) warns against using an over restrictive maximalist definition of bilingualism, such as the "native-like control of two languages" (Bloomfield, 1933, p. 6), as well as an overinclusive minimalist definition, such as Diebold's (1961, p. 99) term *incipient bilingualism* which includes the very first, low-proficiency stages of learning a second language. In this thesis, the definition of a bilingual falls somewhere in between: any person able to understand and use two or more languages. It is still important to provide additional information about the bilinguals' language background and experience in a way that makes it possible to distinguish types of bilingual populations and compare experimental results. This is no trivial matter, and Bloomfield even pointed out that "one cannot define a degree of perfection at which a good foreign speaker becomes a bilingual: the distinction is relative" (1933, p. 6).

Many experimental designs compare groups of bilinguals divided according to certain criteria and/or compare bilinguals to monolinguals. However, since bilingual profiles can vary greatly, it is difficult to establish criteria for grouping and describing bilinguals in a consistent manner across studies. Several factors relating to bilinguals' language background and language experience have been investigated as key variables in experimental studies, such as proficiency (e.g., van

Hell & Tanner, 2012), language dominance (balanced/unbalanced bilinguals) (e.g., Anthony & Blumenfeld, 2019), age of L2 acquisition or arrival in an L2 speaking country (e.g., Flege et al., 1999), amount and type of exposure (e.g., Bonfieni et al., 2019; E.-C. Wu, 2011), language aptitude (e.g., Abrahamsson & Hyltenstam, 2008) and phonological memory (e.g., O'Brien et al., 2007). This shows that there are several sources of individual variation to consider when studying language behaviour. A detailed discussion of these issues is deferred to Chapter 4 of this thesis which reviews available methods for assessment of variation in bilingual profile and proficiency in general, relevant facets of bilingual language experience, and measures relating to the domains of phonology and accent.

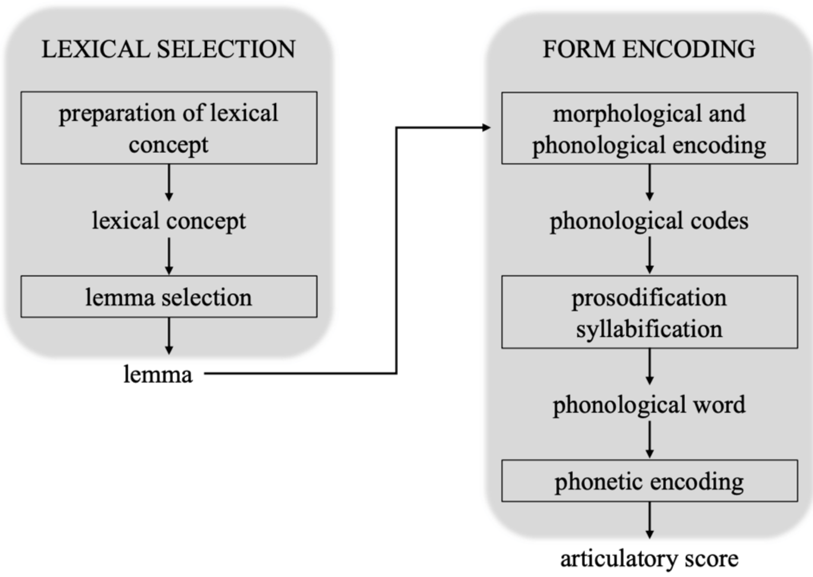
2.3 Word production

In this thesis I use a simple picture naming task to elicit speech and measures of word production. Psycholinguistic models of word production very generally view word production as the serial activation of distinct components that generate the speech output. This section will start by describing the development and basic assumptions of the classic model developed by Levelt (1989), before introducing some alternative mechanisms and how they might accommodate bilingual language production. Word production models were made to account for word production in a stable monolingual system, but there is no comprehensive theory for L2 production so far, as L2 research has mostly focused on certain aspects of word production, such as cross-linguistic influence (Colantoni et al., 2015).

Levelt (1989) described spoken word production in terms of three sequential, specialised components (which may contain subcomponents). The first involving *conceptualising*, where the intended message or idea the speaker wants to convey is prepared by selecting the appropriate concepts and assigning a thematic structure. The output is a preverbal message. Next, the speaker needs to formulate the preverbal message as a linguistic structure. The *formulation* process at the lexical level involves two distinct processes: grammatical and phonological encoding. During formulation, lexical, semantic, and syntactic information are represented separately from phonological information. The process of grammatical encoding involves accessing lemmas, which contain information about the meaning and syntax associated with a lexical item. In addition, there are procedures for generating a syntactic surface structure. Phonological encoding, or form encoding, uses information about an item's lexical form, i.e., morphology and phonology, to generate a phonological and then an articulatory plan. In *articulation*, the articulatory plan is executed by the articulatory system to produce speech (Levelt, 1989).

The model has been revised and developed further by Levelt and colleagues, and the following is based on Levelt (1989, 2001) and Levelt, Roelofs, and Meyer (1999). Here, the conceptualiser is not viewed as strictly autonomous, as it needs to incorporate relevant input that is involved both in preparing and adjusting the conceptual intent of the speaker. It monitors both internal and external speech produced by the speaker, as well as speech produced by others. The other components are viewed as autonomous, and their operation is not affected by processing activity in other components. There is cascading activation from the conceptual level to the lexical level (retrieval of related lemmas) but, after the intended lexical item is selected, there is no activation of unselected representations. In this view, form encoding consists of sequential sub-processes (phonological code retrieval, syllabification, phonetic encoding) ending with the initiation of articulation (visualised in Figure 1).

Figure 1
Diagram of the word production process adapted from Levelt (2001) and Levelt et al. (1999)



Note. Processing stages in boxes, arrows pointing to the output from each processing stage, that serves as input for the next processing stage.

In the above model, processing is largely discrete, apart from at the conceptual level. Activation flows in a top-down manner, with no feedback, and processing is sequential. While there is general agreement that during the first step of conceptualisation, activation of the target representation also activates semantically related concepts (Caramazza, 1997, p. 203; Dell, 1986, p. 291; Levelt, 1989, pp. 183–184), there are different views on the form of activation spreading at subsequent stages. Models assuming a cascade of activation do not contain feedback, but activation flows through the system from all lexical items which have been activated to all connected lemmas and so forth (Caramazza, 1997). Interactive models assume that there is both a cascade of activation and feedback (Dell, 1986, 1988; Dell et al., 1999). Findings of phonological activation from distractor pictures (which are not selected) can be explained by models assuming cascade or interaction (Morsella & Miozzo, 2002, p. 360; Navarrete & Costa, 2005, p. 370; see Wheeldon & Konopka, 2018, for a review).

So far models of monolingual word production have been discussed. The following description of the process of naming a picture in L2 English aims to show how these models are applicable to certain aspects of bilingual picture naming, but that there are complicating factors. For instance, a speaker may be tasked with naming the picture presented in Figure 2. First, they need to identify the object in the picture and understand what it symbolises. At the conceptual level the preverbal message is formed and activates the lemma information associated with "clover". This may for instance include that it is a small plant (meaning) and that it's a noun (syntax). The surface structure then undergoes phonological encoding and, if all has gone according to plan, the speaker will say a word that matches the picture.

While this might seem like a straightforward operation, the influence of the L1 on L2 naming has not been factored in. Beginning at the conceptual level, bilingual naming in their dominant language is found to be slower and more error prone than monolingual naming (Gollan et al., 2005). Tip-of-the-tongue states or problems with retrieving a word are generally more common for bilinguals than monolinguals, however, this is not the case for translatable cognates (Gollan & Acenas, 2004). *Clover* (/ˈkləʊvə/) is a cognate with the Norwegian *kløver*

(/ˈklœʋeɾ/)[2]. If the speaker knows both words, this increases the likelihood that the word will be retrieved and it has a positive effect on the speed of retrieval (Costa et al., 2005). This suggests that conceptual or semantic representations are shared or closely linked between the two languages (Kroll & Stewart, 1994).

Figure 2
Example of stimuli used in the picture naming task

Note. Picture 40 from the MultiPic database (Duñabeitia et al., 2018).

In speech production there seems to be a positive effect of cognates in naming. If this originates at the conceptual level, all types of word production models as described above would be able to account for this finding. However, some authors argue that these facilitatory effects are best accounted for in a model where representations of the non-target language are active at the lexical and phonological level, as well the semantic level (Kroll et al., 2006) and by assuming interactivity between lexical and sublexical levels of processing, both within and across languages (Costa et al., 2005).

At the level of phonological encoding, there is less agreement on whether representations of non-target cognates are activated and on the extent to which there is overlap between languages. The phonological representations at this level, for monolingual production, were said to be abstract, discrete, static and context free. In this sense there is no reason to assume there are different phonological representations. However, phonemes that exist in the L2 but not in the L1, could also be represented in the system. Some speakers even seem to have separate representations (or at least articulation) for sounds that are similar enough to be treated as the same sound (in transcription and by other speakers), but still produce

[2] Transcriptions are based on Standard Southern British English and Eastern Norwegian.

systematic differences when looking at the productions acoustically (E.-C. Wu, 2011). In summary, cognates provide some evidence for shared or closely linked representations between languages, however both the degree of overlap at different levels of representation and the consequences for language processing remain topics for debate.

In this section I have provided an overview of the processes and levels of representation involved in word production. Some differences between bilingual and monolingual word production were discussed, particularly how cognate words compared to noncognate words influence production. However, bilingual word production is not only affected by words that are similar in form and meaning across languages. Research suggests that both languages are activated during bilingual language processing both in perception (De Groot et al., 2000; Y. J. Wu et al., 2013) and production (Colomé & Miozzo, 2010; Gullifer et al., 2013; Lauro & Schwartz, 2017), suggesting that a control mechanism is active during word production, allowing the speaker to produce the intended language.

2.4 Bilingual language control and language switching

Language control has been investigated experimentally through language switching tasks. Asymmetric switch costs (larger costs switching from an L2 into an L1 than vice versa) have been observed for unbalanced bilinguals (Meuter & Allport, 1999), suggesting that an L1 is more strongly suppressed when speaking in an L2 than vice versa. This is consistent with the finding that this asymmetry is attenuated in more balanced bilinguals (Costa & Santesteban, 2004). Switch costs are not only modulated by proficiency, larger switch costs are also found in mandatory cued language switching compared to picture naming where participants voluntarily switch languages, that is, use the first language that comes to mind (Jevtović et al., 2020). When the language change occurs between sentences, resembling code-switching between sentences, bilinguals can switch between languages at no cost (Gullifer et al., 2013). Different models of language control have been proposed.

The inhibitory control model (ICM) accounts for language control in non-selective activation by proposing a number of mechanisms for language control including language task schemas (setting the task of speaking in a specific language) and a checking procedure comparing activated concepts and lemmas to ensure the appropriate language is active (Green, 1998). Activated lemmas that belong to the non-target language are actively inhibited. In this system more active lemmas receive more inhibition. For unbalanced bilinguals, their stronger language is more active and in turn inhibited more strongly than their weaker language, predicting asymmetric switch costs in a switching task.

The adaptive control hypothesis (ACH) describes how different control processes are engaged and adapted to the needs of different communicative contexts (Green & Abutalebi, 2013). In a single language context, where a bilingual only uses one of their languages, language control is achieved by goal maintenance and control or checking mechanisms to ensure the intended language is selected. In a context where two languages are spoken, but not by the same interlocutor, additional control is necessary. Language cues need to be detected and the speakers need to disengage and engage tasks as they are switching between languages, this contributes to switch cost as "the speed of switching from one task to another depends on this disengagement-engagement cycle" (p. 519). In the final context, where bilingual speakers of the same languages can freely switch between languages (i.e., code-switching), similar to the voluntary switching mentioned above, they propose that speakers use the word that is most easily accessible and that there is no need for additional control processes.

There is a general consensus that there is non-selective activation during bilingual language processing (for a non-selective bilingual language comprehension model see Dijkstra and van Heuven, 2002), and that there therefore is a need for a control mechanism that allows the bilingual to speak in the intended language (e.g., Declerck & Philipp, 2015). However, there are different views on for instance how the languages are controlled, e.g. through inhibition (Green & Abutalebi, 2013) or a non-inhibitory competition resolution mechanism (La Heij, 2005). Even though there are unresolved issues, an adequate model of bilingual language control must account for some key findings, such as effects of language production context (e.g., Hanulová et al., 2011), relative language proficiency (e.g., Costa & Santesteban, 2004), and cross-language activation (e.g., Hoshino & Kroll, 2008).

Further evidence for cross-language activation comes from language switching studies manipulating cognate status, however the findings on switch costs and cognate effects are not uniform. Two studies with unbalanced bilinguals found a cognate facilitation effect in both languages and a larger benefit in L1 (Christoffels et al., 2007; Verhoef et al., 2009). Verhoef et al. (2009) additionally manipulated preparation time before naming. After short intervals they found asymmetrical switch costs, however, with longer preparation time they found symmetrical switch costs. The effect of preparation time has implications for the role of inhibition in bilingual language production. In Christoffels et al. (2007) unbalanced bilinguals named pictures with no preparation time and they also found symmetric switch costs, suggesting that whether switch costs are symmetrical or asymmetrical is not only determined by language dominance.

A study with highly proficient bilinguals found symmetrical switch costs and evidence for both cognate facilitation and inhibition (Broersma et al., 2016). The direction of cognate effects was modulated by language dominance. These results all indicate that cross-language activation extends to the phonological level, most often resulting in cognate facilitation. However, the latter study could indicate that cognates also compete for selection at lexical and semantic levels of processing and depending on the strength of facilitation at the phonological level, this could either result in facilitation or inhibition. These findings also highlight the importance of considering individual differences in language processing and language control.

Differences between bilingual and monolingual language processing, especially the additional need for control in bilingual language production, have also been linked to bilingual benefits in non-linguistic cognitive function (e.g., Bialystok & Majumder, 1998; Costa et al., 2008; Prior & Macwhinney, 2010; Woumans et al., 2015; but see Paap et al., 2015; von Bastian et al., 2016), such as the ability to switch between tasks (Prior & Macwhinney, 2010), episodic memory (Schroeder & Marian, 2012), and ignoring conflicting information (Costa et al., 2009). Research also shows that cognitive benefits may be more pronounced in childhood and older age (Bialystok et al., 2012), and that they can be modulated by individual differences, such as language-switching experience (Prior & Gollan, 2011; Verreyt et al., 2016), the age when bilinguals start using both languages actively (Luk et al., 2011) and language dominance (Woumans et al., 2015).

To summarise, there is consistent evidence for non-selective language activation, a need for control mechanisms and cross-language activation at the level of phonology. The experimental effects reported in this section are modulated by individual differences in several aspects of bilingual language proficiency and experience. However, few consider individual variation in the domains of phonology and accent. We return to the issue of individual differences in word production and language switching in Chapter 5 of this thesis.

2.5 Phonological representations and L2 perception and production

In both bilingual word production and language switching, language proficiency differences have been shown to modulate behaviour. Cognate effects have been observed in both types of language production studies, indicating that words sharing form and meaning across languages are represented and processed differently from noncognates. In production, cognate effects are at least partly attributed to the phonological similarity between words (Costa et al., 2005; Hoshino & Kroll, 2008; Verhoef et al., 2009) and it has been argued that bilingual phonological representations are shared between languages (Roelofs & Verhoef, 2006). Studies have also found that articulation (in terms of acoustic characteristics) is affected by cognate status both in word naming (Amengual, 2016a) and language switching (Goldrick et al., 2014). This raises the possibility that how speech sounds are represented and articulated could affect cross-language competition and therefore influence the speed of naming and language switching.

Most accounts of phonological representations view these as normalised, abstract categories made up of key features that help identify and separate a given speech sound from another (e.g., Lahiri & Reetz, 2010; Stevens, 2002). This normalisation or abstraction is often viewed as necessary because the speech signal contains a great deal of variation. An alternative to this abstractionist view comes from accounts based on exemplar theory which assume that detailed collections of heard instances of sounds are stored, rather than one abstract representation for each sound (e.g., Bybee, 2001; Pierrehumbert, 2001). Regardless of the nature of the representations, theories need to account for how the system copes with variation and is attuned to relevant language-specific sound contrasts.

In L1 development, language experience during the first year of life leads to increased sensitivity to language specific sound patterns (Kuhl et al., 2006), while the sensitivity to variations that are specific to other languages decreases (Werker & Tees, 1984), gradually forming language-specific phonological categories. For sequential bilinguals learning an L2, sound perception is affected by the existing L1 system. This is a central assumption in models of L2 perception and speech learning, such as the Speech Learning Model (SLM; Flege, 1995, 2007), the Perceptual Assimilation Model for L2 (PAM-L2; Best & Tyler, 2007), and the L2 Linguistic Perception (L2LP; Escudero, 2005; van Leussen & Escudero, 2015).

Flege's SLM is an influential model describing the process and possible outcomes of L2 speech learning. Initially, L2 speech is processed through the L1 system, but as the listener gains L2 experience, new L2 representations or merged L1-L2 categories may be established depending on the perceived difference between L1 and L2 sounds. In this view, representations exist in a "common phonological space" and can mutually influence each other, both in perception and production. The reviewed cognate effects also suggest shared or closely linked representations at this level of production. It is important to mention that this does not mean that there is a direct link, or complete overlap, between perception and production. Further support for this view comes from neuroimaging studies showing that related brain regions are involved in perception and production (Fridriksson et al., 2009), but that these only partly overlap and are differentially activated depending on the processing task (Hickok & Poeppel, 2000, 2007).

Cognate effects provide evidence of cross-language activation at the phonological level, where language representations appear to be shared or closely linked to some extent. According to the SLM, L1 phonological representations can change over time, and new representations can be formed through L2 exposure. This suggests that the cross-language similarity between cognates is not only language dependent, but also subject to individual variation. In the current project I therefore include self-reported measures from the domains of phonology and accent, and acoustic measures of articulatory divergence, in a study of individual differences in word production and language switching.

3 Motivation and methodology

3.1 Introduction

In this chapter I present the motivation for the current project and areas of enquiry. The experimental components used to address these areas and the general methodology employed are also introduced. This study was designed to investigate individual differences in bilingual word production and uses several types of data, including a bilingual profile questionnaire, objective tests of language proficiency, psycholinguistic experiments that test spoken word production and language switching, and acoustic measures of L1 and L2 divergence in articulation. The results from the experiments are presented in three different chapters. First, the measures of bilingual language profile are analysed and evaluated in Chapter 4. Chapter 5 reports the analysis of the results from the picture naming and language switching tasks, with two bilingual profile factors and performance on a flanker task added as predictors. Chapter 6 details the development of a forced alignment model and the extraction of acoustic measures. From these, articulatory divergence measures were calculated and compared to bilingual profile measures. The divergence measures were also entered as predictors in models of picture naming and language switching.

3.2 Motivation and study description

Decades of research into bilingualism has furthered our understanding of how language processing and performance on linguistic tasks, as well as non-linguistic cognitive tasks, are affected by the presence of two or more active languages and the characteristics of bilingual language profiles. While there is general consensus that both of a bilinguals' languages are activated during language processing, there are different views on the exact nature of this activation, as well as how language selection is controlled. Behavioural evidence demonstrates that both languages are activated both in perception (De Groot et al., 2000; Y. J. Wu et al., 2013) and production (Colomé & Miozzo, 2010; Gullifer et al., 2013), but there are different views on which levels of representation are involved.

The additional demand for control in bilingual language processing, compared to monolingual processing, is often associated with benefits in non-linguistic cognitive functioning (meta-analysis Adesope et al., 2010; review Bialystok et al., 2012). However, there are studies that do not find evidence of a bilingual benefit (e.g., Paap et al., 2017; Paap & Greenberg, 2013; von Bastian et al., 2016). Moreover, across studies, how bilingual profiles are described and assessed differs, and there is variation in experimental findings on bilingual language processing and cognitive function. Studying the effects of individual differences can further our understanding of both (bilingual) language processing and cognitive control, as well as the relationship between the two (de Bruin, 2019; Fricke et al., 2019).

This thesis addresses three key issues related to individual differences. While general bilingual profile is assessed, there is a specific focus on the domains of phonology and accent. First, I investigate the relationships between self-reported proficiency, bilingual profile measures, and performance on objective language tests. This includes variables associated with the domains of phonology and accent. The goal is to assess the degree to which measures in the domains of phonology and accent pattern with other measures of language proficiency, and whether they are influenced by the same levels of bilingual profile (Chapter 4). The second issue investigated, is the effect of individual differences on spoken word production. Two factors derived from the analysis in Chapter 4, *General L2 proficiency* and *L2 accent and interest*, and performance on a flanker task, are used as predictors in models of latencies and accuracy in picture naming and language switching (Chapter 5). Finally, the relationships between individual differences in articulation, based on the articulatory divergence between L1 and L2 vowels, and measures of proficiency and bilingual profile are explored, before the effect of divergence on both naming and switching behaviour is assessed (Chapter 6).

To assess bilingual profile, this thesis employs an adapted version of the validated Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) providing a detailed description of bilingual language proficiency and language experience. Questions relating to proficiency and awareness in the domains of phonology and accent were added to investigate whether these areas of language proficiency contribute to observed differences in language control, and especially cognate effects. The validity of the questionnaire is assessed by comparing self-ratings to performance on four different language tests, two of them measuring aspects of phonological memory, completed both in the participants' L1 and L2. The questionnaire data provide a detailed description of the participants' bilingual profile. The complexity of the dataset is reduced through factor analysis, and the factor scores are used to assess the relationship between bilingual profile and language test performance.

The second issue, the effect of individual differences on spoken word production, is investigated through a picture naming task and a language switching task (Chapter 5). Two well-documented effects in word production, cognate effects (e.g., Costa et al., 2000) and language switching costs (e.g., Costa & Santesteban, 2004), have both been found to be modulated by language proficiency. It has also been suggested that the production benefit observed for cognates in bilingual picture naming is at least partly due to cross-linguistically similar phonological representations, however this similarity may not always be beneficial as language switching studies have found evidence for both cognate facilitation and cognate inhibition (e.g., Broersma et al., 2016). Based on these findings two factors extracted from the questionnaire described above, *General L2 proficiency* and *L2 accent and interest,* as well as performance on a flanker task, were added as predictors in models of data from picture naming and language switching tasks to assess the effects of individual differences on bilingual spoken word production.

In Chapter 6, the focus turns to analysing the final output of the speech production process, specifically the relationship between individual differences in articulation and bilingual language production, as well as their relationship to measures of proficiency and bilingual profile. The study considers two measures of articulatory divergence, based on the degree of overlap/divergence between similar but not identical L1 and L2 vowels. The bilinguals studied in this thesis all started learning their L1 before their L2. In the Speech Learning Model (Flege, 1995, 2007), L1 and L2 speech sound representations exist in a "common phonological space" and they can mutually influence each other, both in perception and production. There are several possible outcomes when learning a new L2 speech sound that is similar but not identical to an L1 speech sound, which again could lead to individual differences in production.

The level of proficiency and performance in the domains of phonetics and phonology is not necessarily related to proficiency in other domains (e.g., Jilka, 2009) and may additionally relate to individual differences in non-linguistic areas such as auditory acuity (Franken et al., 2017) and musicality (Slevc & Miyake, 2006). It is therefore of interest to assess how the articulatory divergence measures relate to other bilingual profile measures and objective tests of phonological processing. The reported effects of cognate status on word production suggest an influence of cross-linguistic phonological similarity, however individual differences in the representation, perception, and production of L2 speech sounds, may lead to individual differences in the degree of cross-linguistic similarity for cognates, and speech sounds in general. Therefore, individual differences in articulatory divergence are entered as predictors in the picture naming and the language switching tasks.

This thesis investigates the effects of individual differences in bilingual profile and articulatory divergence on bilingual word production, including language control. To accomplish this, a multi-component study was conducted. The classic psycholinguistic paradigms of picture naming and language switching are employed to assess language production and language control with a new population of bilinguals. Detailed research questions and theoretical reviews are given within each relevant experimental chapter (4-6). The experimental components and general procedure are described below.

3.3 General methodology

The effects of individual differences in bilingual profile and articulatory divergence on bilingual word production were investigated through a series of different tasks: a language experience and proficiency questionnaire, four objective language tests, a simple picture naming task, a picture naming task with language switching and a flanker task. A full overview of the tasks can be seen in Figure 3. Finally, speech collected during the picture naming tasks was analysed acoustically. Descriptions and motivation for each component is briefly summarised below before details of the general procedure are described. The detailed methodology for each component will be described in the relevant experimental chapters of this thesis, indicated in parentheses below. Data collection occurred over three days of testing, and these were conducted within five days for each participant. Each participant completed all single language experiments (in English or Norwegian) in one day. Half of the participants started with English, and the other half started with Norwegian. On the third day, all participants completed the language switching task, the flanker task, and the augmented LEAP-Q.

3.3.1 Experimental components

Language experience and proficiency questionnaire

The questionnaire (adapted from Marian et al., 2007) elicited self-ratings of language proficiency in different domains, amount of language exposure and language use in different contexts (e.g., with family or at work) both current and during language development, time exposed to different kinds of language input (e.g., conversational, reading or media), language dominance, and questions relating to phonology and accent. The responses provided descriptive data of the bilingual participants, and input data for a factor analysis. (Chapter 4)

Objective language tests

Four language tests were completed in both L1 and L2. A spelling task and a vocabulary task served as objective measures of proficiency, while an elision task and a serial nonword recognition (SNWR) task served as measures of phonological awareness, working memory, and phonological short-term memory. Performance was evaluated in terms of accuracy in all tasks. Test scores are used to evaluate the reliability of self-reported proficiency measures. The majority of the tests were constructed for this experiment to create comparable tests in Norwegian and English. For the spelling and the vocabulary tasks, cognate words were excluded to avoid confounding measures of English and Norwegian proficiency.

In the spelling task participants heard and then typed 20 words. In the vocabulary task participants identified which word out of four possible options matched a target word. First, they identified synonyms/near-synonyms (20 target words) and then they identified antonyms/near-antonyms (20 target words). In the elision task, participants first verbally repeated a nonword that they had just heard and then they were instructed to repeat the word again leaving out one of the segments, measuring their ability to retain and manipulate nonwords in working memory (Sasisekaran & Byrd, 2013). In the SNWR task, participants needed to retain and compare two auditorily presented sequences of nonsense syllables. Syllables were either presented in the same order both times (same trials) or two adjacent syllables swapped position in the second presentation (different trials). This served as a measure of phonological short-term memory. (Chapter 4)

Figure 3
Overview of experimental sessions

Day 1 - single language

Information and consent form
Colour vision check

Task type	Duration	Components
Picture naming	40 minutes	Familiarisation
		Practice block
		Experimental blocks
Reading	15 minutes	Single words
		Memorised sentences
	BREAK	
Language tests	25 minutes	Elision
		Spelling
		Serial nonword recognition
		Vocabulary

Day 2 - single language

Task type	Duration	Components
Picture naming	40 minutes	Familiarisation
		Practice block
		Experimental blocks
Reading	15 minutes	Single words
		Memorised sentences
	BREAK	
Language tests	25 minutes	Elision
		Spelling
		Serial nonword recognition
		Vocabulary

Language counterbalanced - either L1 - L2 or L2 - L1

Day 3 - both languages

Task type	Duration	Components
Switching task	25 minutes	Practice block
		Experimental blocks
Flanker task	10 minutes	Practice block
		Experimental blocks
Questionnaire	20 minutes	
Experiment debrief	5 minutes	

Bilingual word production and language control

Bilingual word production and language control was assessed by collecting reaction times and error rates in picture naming experiments, which were completed both in a single language and a switched language context. Participants were instructed to name pictures as quickly and accurately as they could. In the simple picture naming task, all pictures were named in L1 and L2 on separate days. A subset of the pictures was used in the language switching task. A coloured frame around the picture indicated which language they should use. Finally, a version of the flanker task (Zhou & Krott, 2018) served as a non-linguistic measure of selective attention and control. (Chapter 5)

Speech collection and acoustic analysis

Speech was collected during the picture naming and language switching tasks[3]. Two L2 English vowels that tend to be challenging for many proficient native speakers of Norwegian, and two L1 Norwegian vowels that speakers tend to use in place of the English vowels were selected from the speech materials. The acoustic difference between L1 /ʉː/ and L2 /uː/, and between L1 /œ/ and L2 /ʌ/, was measured to quantify articulatory divergence. The resulting divergence measures were then compared to self-reported accent proficiency, language test results and factors found in Chapter 4. Finally, the articulatory divergence measures were entered as predictors into models of bilingual language production and control. (Chapter 6)

3.3.2 General procedure

Participants

63 participants were recruited through advertisements on campus. The inclusion criteria were that participants should have Norwegian as their L1 and English as their strongest L2. They should be aged between 18-35, have no diagnosed language difficulties, and normal (or corrected to normal) vision and hearing. Informed consent was obtained prior to the experiments, and they received a gift card for their participation. A copy of the information sheet and consent form is available in Appendix A.

[3] In addition, participants were recorded reading words and repeating memorised sentences, but this speech was not analysed for this thesis.

All participants grew up with Norwegian as a home language, and some with both Norwegian and English spoken in the home. Two participants were hindered from completing all experimental sessions within five days and their data were excluded from the study. One participant was excluded because the dominant L2 was not English, reducing the number to 60 participants. One participant reported that Norwegian and English dominance was completely balanced. Proficiency and test scores were not outside the normal range of responses and this participant was retained for the analysis.

Apparatus

All participants were tested individually and all experiments were conducted in a sound insulated booth. Sennheiser GSP 350 headphones with a noise-cancelling microphone were used to record participant responses and to play auditory stimuli. Sound stimuli were played at a comfortable listening level, and participants could adjust the sound level if necessary. The headphones were worn by the participants in all three sessions. Responses were collected by voice key or keypress, depending on the type of experiment. Participants were seated approximately 75 cm from a 23-inch iiyama screen with a 1920 x 1200 resolution.

The picture naming task, switching task, sentence reading task, and word reading task were built in Presentation (version 20.1, Build 12.04.17). All audio was recorded in stereo at a 48000 Hz sampling rate. The spelling task, vocabulary task, elision task, serial nonword recognition task, and flanker task were built in OpenSesame version 3.1.9 (Mathôt et al., 2012). The OpenSesame screen resolution was 1146 x 798.

3.4 Ethical approval

The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation. A note on ethical considerations for this project can be found in Appendix B.

4 Language background and proficiency

4.1 Introduction

As described in Section 2.2, the term bilingual, used in this thesis, refers to any person able to understand and use two or more languages. However, there are several ways of defining, describing, and grouping bilinguals. This is not surprising since language acquisition, exposure, and use will vary across languages, cultures, and individuals. Research has also identified several aspects of the bilingual language experience thought to influence both language-related (e.g., Bonfieni et al., 2019; Love et al., 2003) and non-linguistic cognitive processes (e.g., Martin-Rhee & Bialystok, 2008; Prior & Macwhinney, 2010; Verreyt et al., 2016). In addition, language proficiency has proved to be a powerful predictor of language behaviour (e.g., Anthony & Blumenfeld, 2019; Costa & Santesteban, 2004; Rosselli et al., 2014).

In this project the complex nature of bilinguals and their language processing is studied by collecting data on bilingual profile, including language proficiency, and language behaviour. This chapter reports the development of, and results from, eight tests and a questionnaire (experiments addressed in this chapter are framed in Figure 4). Four language tests: vocabulary, spelling, serial nonword recognition (SNWR) and elision, were all conducted in both Norwegian and English. The questionnaire was an augmented version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007). Both the original LEAP-Q and the augmented questionnaire are provided in Appendix C. The main purpose of the augmentations (described in detail in Section 4.3.1 and Appendix C) was to obtain more information relating to accent, phonology, and language switching. The language tasks serve as objective measures of proficiency (vocabulary and spelling) and measures of phonological processing and phonological memory (elision and SNWR).

This chapter will first provide an overview of how different types of bilingualism, language proficiency and language experience have been defined and assessed in previous studies of bilingual language processing. Then the methods, including design and materials, used in the current study will be described and finally the analysis and results relating to language background and proficiency will be presented.

Figure 4
Experiment overview Chapter 4

Day 1 - single language	Day 2 - single language	Day 3 - both languages
Picture naming	Picture naming	Switching task
Reading	Reading	Flanker task
Elision	Elision	Questionnaire
Spelling	Spelling	Experiment debrief
Serial nonword recognition	Serial nonword recogntition	
Vocabulary	Vocabulary	

Language counterbalanced - either L1 - L2 or L2 - L1

Note. Experiments addressed in Chapter 4 are framed.

Two research questions relating to self-ratings in the domains of phonology and accent are addressed in this chapter:

- What is the relationship between self-ratings in the domains of phonology and accent and self-ratings of other areas of bilingual profile?
- Is there a relationship between self-ratings in the domains of phonology and accent and objective measures of phonological processing?

A factor analysis of the questionnaire responses suggests that the inclusion of phonology and accent questions in the questionnaire contributes to the assessment of bilingual language profile. These questions grouped together in a factor that was named *L2 accent and interest*. However, there is some overlap with general language proficiency and language exposure. Language tests and self-reported measures were generally more related for L2 than L1. Of the phonological tests, elision scores, but not SNWR scores, correlated with self-ratings of pronunciation proficiency in both languages. The *L2 accent and interest* factor correlated with L1 elision, but none of the other phonological tests.

4.2 Theoretical overview

In this thesis a bilingual is defined as "any person able to understand and use two or more languages", but as discussed in the theoretical overview (Chapter 2), bilingual processing is influenced by several variables relating to language background and language experience. It is therefore necessary to describe bilingual participants in experimental studies in greater detail. In this section I first describe different ways of assessing and defining language proficiency, before focusing on aspects of bilingual profiles relating to language experience and language background. The last section will discuss ways of defining and assessing individual differences and proficiency related to the domains of phonology and accent.

4.2.1 Assessing and defining language proficiency

Many studies compare different groups of bilinguals to each other and/or monolinguals. Bilingual group affiliation is often decided by some measure of language proficiency, either proficiency level (such as high, intermediate, or low) or language dominance (balanced or non-balanced bilinguals). Many language phenomena are modulated by the proficiency of the speaker, such as asymmetrical switching costs (e.g., Costa & Santesteban, 2004) and the facilitatory effect of cognates (e.g., Costa et al., 2000), which have been found to be reduced or not observed for highly proficient or balanced bilinguals.

There are several ways of assessing language proficiency. The most common methods include placement tests, self-assessment, and (standardised) domain specific tests. Studies may report one or several measures of proficiency. In cases where English is the language being studied, proficiency may be based on general scores from tests such as the Test of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS). The Common European Framework of Reference for Languages (CEFR) may also be used in a European setting, or enrolment in the same level of language course. These measures are perhaps most suitable for group-wise comparisons of native speakers compared to L2 speakers of the same language and/or different levels of L2 proficiency.

Self-ratings of proficiency or self-assessments of language skills are frequently used both in studies focusing on language processing and language learning (e.g., Bernolet et al., 2013; Flege et al., 2002; W. Ma & Winke, 2019; Wharton, 2000). Depending on the purpose of the study, a few or several self-ratings may be obtained. In a meta-analysis of the relationship between self-rated and tested language proficiency, Ross (1998) analysed 60 correlations reported between various measures of language proficiency and self-ratings. An overall significant correlation was found between measures of language proficiency and test results, but the author also mentions that there was a great deal of variation within the sample and not all language domains show the same degree of correlation, suggesting that certain aspects of proficiency may be easier to self-rate than others.

The reliability of self-ratings may also be affected by whether the L1 or L2 is being rated, but here findings are unclear. For instance, Delgado et al. (1999) conducted a study where Spanish-English bilinguals, primarily L1 Spanish, rated their own L1 and L2 skills before and after they completed a series of language tests. The results showed a correlation between self-ratings and test scores for all Spanish tests, but for the English tests less than half of the correlations were significant, suggesting ratings were more reliable for L1. However, other studies report strong correlations between self-ratings and tests for both languages, but with stronger correlations for L2 than for L1 (Marian et al., 2007).

While studies generally find self-ratings to be reliable, research also suggests that individual differences influence participant responses. For instance, the degree to which participants experience anxiety associated with speaking in their second language may affect the reliability of self-ratings. It has been found that more anxiety is correlated with underestimating language competence, while less anxiety is correlated with overestimating competence (MacIntyre et al., 1997). Similar effects have been found for proficiency, where less proficient speakers tend to overestimate their performance, while more proficient speakers tend to underestimate their performance (Trofimovich et al., 2016). Self-ratings of proficiency level may also vary depending on aspects of bilingual profile, such as language dominance, and the languages spoken by the bilingual (Tomoschuk et al., 2019). For these reasons it has been argued that (standardised) objective measures of proficiency should be included in the assessment of bilingual language proficiency (e.g., de Bruin, 2019; Tomoschuk et al., 2019).

Several standardised objective measures of language proficiency are also available, albeit not for all languages. The Peabody Picture Vocabulary Test (PPVT) is a standardised test measuring receptive vocabulary and it is also used as a measure reflecting general language proficiency in some studies (e.g., Luk et al., 2011). The Boston Naming Test is an example of another test that was not developed specifically for bilinguals, but that has been used in studies with bilingual participants. This test measures productive vocabulary and has for instance been used for investigating cognate effects (Rosselli et al., 2014) and for assessing language dominance (Hernandez & Meschyan, 2006). Tests developed specifically for bilinguals include The Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012), measuring vocabulary knowledge and proficiency, and the Multilingual Naming Test (MINT), measuring productive vocabulary, and also considering cross-language similarity between words and difficulty when selecting items to be named (Gollan et al., 2012).

These tests are often used in experiments to determine language proficiency and/or language dominance. Challenges associated with measuring language proficiency in bilingualism research is addressed by Hulstijn (2012). This paper reviews how language proficiency can be measured and defined, especially regarding selecting a suitable monolingual control group for the bilingual group being studied, but these concerns are also valid when comparing bilinguals. One important consideration when creating or selecting language proficiency tests, according to Hulstijn, is being aware that language tests measuring the same skill (e.g., spelling or vocabulary) will not be directly comparable for two different languages and can often be likened to "comparing apples and oranges" (2012, p. 427). This is relevant for the current study, for instance when assessing spelling, since Norwegian orthography is more transparent than English orthography (Seymour et al., 2003).

In a review paper, de Bruin (2019) discusses methods for assessing bilingual language proficiency and language experience, and recommends providing both detailed descriptions of bilingual profile in addition to objective standardised measurements. However, the author points out that standardised tests are subject to different availability in different languages. This was the case in the current study, where no standardised language tests that were both suitable for young adults and available in both Norwegian and English were identified. Therefore, Norwegian and English language tests were developed for the current study. The paired language tests were identical in structure, and based on existing tests, but the test items were selected to challenge proficient young adults and by focusing on typical errors and challenges particular to each language.

4.2.2 Language experience and language background

A lot of research on bilingualism has compared groups, for instance categorised by proficiency or age of acquisition. In addition to proficiency, different types of bilingualism are also associated with differences in language experience and language background. Many researchers include supplementary information, such as the participants' self-reported age of acquisition, years of formal language instruction and previous and current language experience (e.g., Dijkstra et al., 1999; Duyck et al., 2007), but the level of detail, and whether these variables are entered into analysis varies.

The Language Experience and Proficiency Questionnaire was built to reliably assess bilingual profiles (LEAP-Q; Marian et al., 2007), where bilingual profile, as the title suggests, consists of both language proficiency measures and factors relating to the participants language experience, such as the age of language acquisition, language learning context and current language use. Other questionnaires have also been developed to provide a more thorough assessment of language experience and language background in addition to proficiency ratings. Such as the Language History Questionnaire developed by P. Li et al., which is an online questionnaire based on the most frequently asked questions in language questionnaires (P. Li et al., 2006, 2014, 2019) and the Language and Social Background Questionnaire (Anderson et al., 2018). Detailed language questionnaires provide a great deal of data about participants, but the number of variables often need to be reduced when using this data for further analysis. Different options for data reduction have been employed, such as factor analysis and language entropy. Gullifer and Titone (2020) introduce language entropy as a continuous individual difference measure of the relative balance between bilinguals daily use of their languages based on questionnaire responses. Several studies have used factor analysis for deriving underlying constructs that capture the main sources of individual variation and/or to assess the validity of questionnaires (e.g., Anderson et al., 2018; Luk & Bialystok, 2013; Marian et al., 2007).

4.2.3 Assessing spoken proficiency, phonology, and phonetic production

Assessing proficiency in the speech domain, particularly related to accent, phonology, and phonetic production, is of special interest to the current study. Measures of L2 oral proficiency, generally focus on fluency, comprehensibility and accentedness. Fluency is typically assessed in longer stretches of speech and objective measures include speech rate, mean length of fluent stretches of speech, number of corrections and repetitions, and number and duration of both silent and filled pauses (De Jong et al., 2013; O'Brien et al., 2007). Comprehensibility is typically assessed by listener ratings of how easily the speech is understood (e.g., Saito et al., 2017; Trofimovich & Isaacs, 2012). Comprehensibility appears to be partially related to accent, but a study found that comprehensibility was more strongly affected by the grammatical and lexical accuracy of the speech, while accent was related to variables associated with phonology, such as accuracy at the syllable and segment level (Trofimovich & Isaacs, 2012).

Listener characteristics can influence ratings of the accentedness of L2 speech. Studies have for instance found that the perceived degree of L2 accent is influenced by the listener's degree of familiarity with L2 accented speech (Schmid & Hopp, 2014) and the range of accentedness in the sample as a whole (Flege & Fletcher, 1992). Listener strategies may also affect ratings. In a study where 10 native listeners rated the overall degree of perceived foreign accent in L2 English, one of ten listeners used the whole scale, while the others avoided either the lower or higher end of the scale all together (Flege et al., 1995).

Self-ratings of L2 accent and comprehensibility can also be affected by L2 proficiency. Trofimovich et al. (2016) compared L2 self-ratings of accentedness and comprehensibility to native listener ratings of recorded L2 speech from the same participants. They found that L2 participants with lower native listener ratings, i.e., whose speech was rated as less comprehensible and more accented, tended to overestimate their own performance, while L2 participants with higher native listener ratings tended to underestimate their own performance. The participants associated with different levels of L2 accent and comprehensibility also had different L1 backgrounds, which may influence both native ratings and self-ratings. The results suggest that depending on the L1, participants may focus on different aspects of L2 production when rating their own performance. A study comparing L2 self-ratings of accuracy on individual speech sounds to native ratings of the same L2 productions found that the two groups agreed on 85% of the ratings overall (Dlaska & Krekeler, 2008). However, while L2 participants and native listeners generally agreed on which L2 sounds were correctly produced (89%), agreement was lower for inaccurate sounds (44%). The authors discuss factors that could have contributed to the differences between native listener ratings and L2 self-ratings, such as influences from L1 phonology which may make it difficult for the L2 participants to perceive and produce L2 contrasts.

Ratings of proficiency are commonly used for assessing bilinguals in general, as described earlier in Section 4.2.1, and both self-ratings and native ratings are also used for assessing proficiency in the speech domain. Some potential sources of bias when using ratings for evaluating L2 speech were pointed out above. One way of avoiding these biases is by using objective measures focusing on quantifiable characteristics of speech produced by bilinguals. The start of this section mentioned three aspects of L2 speech that contribute to proficiency in this domain, namely fluency, comprehensibility and accentedness. The focus of this thesis is investigating the effects of individual differences in bilingual profile and articulatory divergence on bilingual word production. Therefore, the influence of individual differences in phonology and accentedness on the word level and segment level is of greater interest than measures of fluency and comprehensibility.

Accent has been found to be related to variables associated with accuracy at the syllable and segment level (Trofimovich & Isaacs, 2012). A combination of factors can contribute to the perceived accentedness of speech, or accuracy in pronunciation, such as stress placement and the quality of spoken syllables and segments. Accuracy on the segment level can be assessed by comparing acoustic characteristics of speech sounds, such as voice onset time (e.g., Antoniou et al., 2010) and vowel formants (Kartushina & Frauenfelder, 2014), produced by bilinguals to the same characteristics measured in speech sounds produced by monolingual speakers. There are also several individual difference factors that may affect a speaker's ability to acquire and produce L2 speech. Proficiency and abilities in the domains of phonology and L2 accent is for instance affected by factors relating to a speaker's bilingual profile, such as age at the onset of L2 learning (Flege et al., 1995) and amount of L2 exposure (Flege, 2018), and individual differences in cognitive skills, such as musical ability (Slevc & Miyake, 2006), auditory selective attention (Mora & Mora-Plaza, 2019), phonological short term memory (MacKay et al., 2001) and phonemic coding ability (Saito, 2017).

Several factors contribute to both accent in L2 production and the perception of accentedness. Proficiency in this domain can be assessed in multiple ways, and the current study will employ a selection of measures, both self-reported and objective, that target different aspects of accent and are suitable for looking at individual differences. The goal is to assess the degree to which measures in the domains of phonology and accent pattern with other measures of language proficiency, and whether they are influenced by the same levels of bilingual profile. The questionnaire includes self-ratings of pronunciation proficiency and questions addressing phonological awareness and interest in the accent domain. In the current chapter the relationship between self-reported variables describing bilingual profile obtained from the questionnaire is first evaluated through factor analysis. Then the results from objective language tests, including measures of phonological short-term memory and phonological awareness, are evaluated. The test scores are compared to self-rated proficiency in relevant domains and to the factors extracted from the questionnaire. General language proficiency has been found to modulate both switching (Meuter & Allport, 1999) and cognate (Costa et al., 2000) effects in bilingual language production. Individual difference measures from the current chapter will be analysed with picture naming and language switching data in Chapter 5. In Chapter 6, the analysis is extended to include individual difference measures of articulation.

4.3 Materials and design

In this project language background and proficiency are assessed by self-reported data and paired L1-L2 language tests. The four behavioural tasks (elision, spelling, serial nonword recognition (SNWR), and vocabulary) were completed in English and Norwegian on separate days. The augmented version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) was completed on the last day to avoid biasing performance. In this chapter the design and procedure of each experimental component will be described, starting with the questionnaire, then the spelling task, vocabulary task, elision task, and finally SNWR.

4.3.1 Questionnaire

Design

The questionnaire is based on the LEAP-Q, which is a validated questionnaire for assessing both monolingual and bilingual language profiles. It elicits self-reported information about language exposure and language history, as well as self-ratings of proficiency. The LEAP-Q and the adapted questionnaire, including all changes from the original, are available in Appendix C.

For this study the LEAP-Q was altered in a number of ways. Questions about dialect, accent, language learning environment and language switching were added, as well as a section on opinions about language learning (adapted from Flege et al., 1999). These changes were made to obtain more information about language experience, and more information about the domains of phonology and accent. Second, three questions were removed from the LEAP-Q because they were asked in a different way later (Question 4), not applicable to the participant group (Question 8), or less relevant for the current study (a part of question 3 on page 2: "Select proficiency in understanding spoken language"). In addition, self-ratings of proficiency were obtained in more domains (accent, writing, grammar, vocabulary, and spelling). Some changes were made to the wording, structure, and question order so that screening questions were grouped, and ratings of proficiency and exposure were filled out for both languages at the same time. Finally, some changes were made to make the questions more current (e.g., replacing *language tapes* with *language learning apps*).

Procedure

When filling out the questionnaire the participant and the experimenter were seated next to each other in front of the computer screen. The experimenter typed participant responses into a spreadsheet. Participants were encouraged to ask the experimenter if any of the questions were unclear. On the first day participants filled out the screening questions in the questionnaire and the rest were filled in at the end of the third day.

To increase the sample size for the factor analysis, 100 additional responses were collected. A version of the questionnaire suitable for printing was created and distributed on campus by a research assistant. The research assistant was available to answer questions the respondents might have while filling out the questionnaire.

4.3.2 Spelling task

Design and materials

The materials consisted of 20 words each in the English and Norwegian version of this experiment (spelling task materials in Appendix D). Eleven of the English words came from a list previously used in a master project in our lab. Nine cognates were removed from the previous experiment and replaced with noncognate words. The list was compiled to reflect the lack of transparency between English orthography and pronunciation and typical errors (e.g., *thoroughly*).

Norwegian orthography is often more transparent than English and finding words with a similar level of difficulty in Norwegian and English is therefore not straightforward. Instead of attempting to create similar lists, Norwegian Bokmål words were selected after gathering examples of common spelling mistakes and identifying problematic areas. Many problems are related to geminates and/or compounds, i.e., erroneous inclusion or omission of a consonant (e.g., *anerkjenne*). Teachers and lecturers were informally asked to give feedback on a prepared list of words and encouraged to suggest other problematic words. The final list was selected seeking a balanced distribution of vowel and consonant errors, and omission and inclusion errors.

Cross-language comparisons of the number of syllables, orthographic characters, and word-frequency all yielded non-significant results (but note that cross-language frequency measures need to be interpreted with caution, see discussion in Section 5.3.1). All participants heard the words in the same order.

The English stimuli were recorded by a native speaker with a Southern British accent at a 44100 Hz sampling rate using a Sony ICD-PX370 digital voice recorder. The Norwegian stimuli were recorded in the same location and with the same specifications as the speech collected from the participants, using Audacity® (Audacity Team, 2019). The speech was then resampled to 44100 Hz and normalised to match the English recordings[4]. The native Norwegian speaker spoke with a Kristiansand dialect, which was familiar to all the participants. For both

[4] Resampled using Reetz' PreProcess script for Praat (Reetz, 2022).

speakers, the items were read from a list with dummy items at the beginning and the end of the list. Stimuli were extracted individually, preserving 300 ms of silence before and after the word.

Procedure

For the spelling task an audio file containing the word to be spelled was played over the headphones. The participants could press a button to hear the word again. They used the keyboard to type in their response, which was visible to them on the screen, and pressed enter to submit their response. The next sound file was played 300 ms after the participant gave their response. Error coding was automatic and case insensitive, and error rates were exported for analysis.

4.3.3 Vocabulary task

Design and materials

Both the English and Norwegian vocabulary depth task consisted of two parts, one identifying synonyms/near-synonyms and the other identifying antonyms/near-antonyms. The target stimuli were 20 low-frequency words in each part of the test and there were four foils for each target stimulus (vocabulary task materials in Appendix D). The stimuli were presented with a different randomisation for each participant and were presented as 24-pixel black text on a white background. The test structure and some of the English words were taken from a test developed by S. Frisson (personal communication, 2018). Since only noncognate items could be used, some words from the original English test had to be replaced. Of the four foils one word would be the correct response (e.g., English synonym: LOQUACIOUS (target), *talkative* (correct) – *broad* – *roomy* – *marshy*; Norwegian antonym: ARMOD (target), *rikdom* (correct) – *avsporing* – *elendighet* - *bopel*). Several strategies were used to create the other foils, for instance using words that were semantically related to the correct response, semantically related to the antonym of the correct response or words that a participant might select if they are guessing and do not know the actual meaning of the word (such as the foil *contemplative* for the target word *ponderous*).

Procedure

For the vocabulary task, the experimenter was seated next to the participant during the experiment and typed in their response. First, participants completed 20 trials where they had to select the synonym of the target word from the foils and then 20 trials selecting the antonym. For each trial, a target word was displayed at the top of the screen and four possible answers, numbered 1-4, were presented below as well as an “5 I don’t know” option at the very bottom. The participants were instructed to choose the word with the most similar meaning to the target word for the synonym task and the most dissimilar word for the antonym task. They were told that in this task some words were very difficult and that they should choose option 5 if they did not know the answer. However, if they had a hunch about which word might be correct, or could make an educated guess, they could still choose this word even if they were not entirely sure that this was the right option. They gave their response by saying the number of the word out loud and the next trial was presented immediately after a response was given. Errors were registered automatically and exported for further analysis.

4.3.4 Elision task

Design and materials

The English and Norwegian versions of this task both included 36 nonwords each. The English version had previously been used for a master’s project in our lab. The Norwegian nonwords were created to match the English stimuli on word length, consonant cluster length and position, and location of manipulation. Both Norwegian and English nonword stimuli followed the phonotactic constraints of the respective language (full stimuli lists in Appendix D). Stimuli were presented in the same order for all participants.

The English stimuli were recorded by a native English speaker with a Southern British accent, in a sound-attenuated room using a professional quality USB microphone (Røde NT-USB) at a sampling rate of 44100 Hz. The Norwegian stimuli were recorded with the same specifications and speaker as for the spelling task. The nonwords were uttered in the context “Say [nonword]” (Norwegian “Si [nonword]”) and words to be manipulated were uttered in the context “Now say [nonword] without the [segment to be removed]” (Norwegian “Si [nonword] uten [segment to be removed]”).

Procedure

At the start of a trial a sound file was played, and the participants were asked to repeat the nonsense word they heard, e.g., “Say /ˈsplɔɪtəl/” (English) or “Si /ˈkɑŋtʁɑk/” (Norwegian). The next sound file instructed participants to remove a phoneme and then repeat the word, e.g., “Now say /ˈsplɔɪtəl/ without the /p/” or “Si /ˈkɑŋtʁɑk/ uten /t/”. Each sound file was played once. Between each trial the experimenter entered whether the response was correct or not and the next sound file was presented after 500 ms. Responses were recorded in mono at a 22050 Hz sampling rate. Error rates were exported and analysed. In general, correct trials contained the correct segments in the correct order. However, if the vowel quality was different or /m/ and /n/ were confused, items were still marked as correct if the order and number of segments matched the instructions.

4.3.5 Serial nonword recognition (SNWR)

Design and materials

For this task participants heard two (increasingly long) sequences of nonsense syllables in each trial. The syllable order was either the same in both sequences (e.g., /pim tɑɹg gæb bɑk tʃɛl/ - /pim tɑɹg gæb bɑk tʃɛl/) or different (e.g., /pim tɑɹg gæb bɑk tʃɛl/ - /pim gæb tɑɹg bɑk tʃɛl/). On *different* trials two adjacent syllables were transposed. The two first and two last syllables of a string were not transposed on critical trials. The task and English stimuli were similar to previous experiments (for instance, Gathercole et al., 2001; O’Brien et al., 2006). For both languages 144 syllables were used to create the experimental stimuli sequences of five to seven syllables (full stimuli lists in Appendix D). Norwegian nonwords were created allowing for a complex onset and seeking a balanced number of occurrences of initial consonant (9 or 10 of each segment type), medial vowel (15-17 of each segment type), medial diphthongs (6-8 of each segment type), and final consonant (9-12 of each segment type). The syllable sequences were assembled so that each syllable within the sequence had a different vowel quality and as few consonant repetitions as possible. The stimuli were pre-recorded by native speakers, the English speaker was the same as in the spelling task and the Norwegian speaker was the same as in the spelling and elision task.

In the current study, the SNWR list was half the length compared to similar studies to prevent learning and fatigue. The syllable sequences were used to create two lists, with half of the stimuli in each list. Half of the participants received one list and half the other. The originally paired strings (i.e., the same syllables used in *same* and *different* trials) were put in separate lists to limit exposure to the strings. The lists were pseudorandomised with the following conditions: 1) no more than three consecutive trials of the same type, 2) no consecutive trials with the same syllable switch location, and 3) the same amount of *same* and *different* trials in each list. Three dummy items, where the last two syllables were transposed, were added to each list to prevent that participants would learn to ignore the syllables in these positions. The three first trials served as practice trials.

Procedure

Two strings of syllables, separated by a 750 ms silent interval, were played over the headphones. The participants had to indicate whether the order of the syllables in the two strings was the same or different. At the offset of the sound file a response screen was shown reminding them to press 1 for "same" and 2 for "different" ("lik" and "ulik" in Norwegian). The next trial began 1000 ms after they gave their response. Each stimulus was played only once. Errors were registered automatically, and error rates were exported for analysis.

4.4 Results

This section reports the results from analysis of the questionnaire data and language experiments. Section 4.4.1 summarises LEAP-Q responses, both for the extended sample responding to the questionnaire (182 participants) and the 60 participants who completed all experiments that were part of this PhD project. In Section 4.4.2 the results from a factor analysis, conducted to look at how questionnaire variables group together and to reduce complexity, is reported. The next two sections report results from the 60 participants in the experimental group. Section 4.4.3 reports results from the four objective language tests conducted in both English and Norwegian. First, the tests themselves are evaluated and then test scores are compared to self-ratings relating in the same domain. Finally, in Section 4.4.4 the relationship between test scores and factors extracted from the questionnaire are investigated.

4.4.1 Questionnaire - descriptive results

Participants

The augmented LEAP-Q was completed by 182 native Norwegian speakers with English as their strongest L2 (132 women, 50 men). All participants reported having some experience with at least a third language, as a second foreign language is compulsory in Norwegian schools. Eighty-three questionnaires were collected during experiments in the Experimental Linguistics Lab at the University of

Agder, 60 from the current study and 23 from an MA study that was run at the same time. The final 100 (99) responses came from questionnaires distributed in paper format on campus by a research assistant who also digitised their responses. Further information about inclusion criteria can be found in Section 3.3.2. The 83 participants who participated in lab experiments, in addition to completing the questionnaire, received gift cards. The participants who just completed the questionnaire on paper graciously volunteered and were not reimbursed. The summary of age, education, and gender in Table 1 shows that the group demographics are similar.

Table 1
Summary of participant age, education and gender grouped by source of data collection

Group	Age *M (SD)*	Age *Range*	Education (years) *M (SD)*	Education (years) *Range*	Male	Female	Non-binary
Exp	24.10 (4.25)	18-35	16.35 (2.33)	12-22	16	44	0
MA	23.57 (3.91)	19-36	16.26 (1.89)	14-20	4	19	0
Handout	22.66 (3.05)	18-34	15.97 (1.80)	13-20	30	69	0
Overall	23.25 (3.64)	18-36	16.13 (2.00)	12-22	50	132	0

Note. From the first row, results are grouped by participants in the experimental group (Exp), participants from a master project in our lab (MA), and participants who answered via the handout questionnaires on campus. The overall summary for all participants is on the last row.

From the handouts, the data from one participant were excluded because of 19 missing responses. There were no missing values in the questionnaires collected in the lab. In the remaining handouts, there were 13 data points missing in total and these values were replaced by the mean of the other responses to the question. Details are provided in Appendix C, Table C1.

LEAP-Q descriptive results

In general, the responses suggest that the participants generally were more exposed to Norwegian than English, and that both learning and exposure to each language to some extent comes from different sources and environments. In the following, the overall results and the results from the experimental group will be described and compared. First, results relating to language learning and language exposure will be discussed before we turn to proficiency ratings and questions relating to accent and dialect.

Table 2 provides a summary of responses to questions associated with language exposure for all participants and Table 3 shows the same data for the experimental group only. On average both groups of participants reached language milestones in Norwegian before English and have been immersed in a Norwegian language environment for most of their life. The participants were asked to rate how much certain types of language exposure contributed to their learning of each language. Reading and education received similar ratings in both languages. For Norwegian, interacting with friends and family generally received higher ratings than for English. This pattern was reversed for TV/streaming, music/media, and self-instruction. While the general pattern is the same in both groups, the range of answers suggests that some questions were interpreted differently by the respondents in the overall group compared to the experimental group.

Participants were asked to rate the total amount of exposure to each language. On average both groups reported being more exposed to Norwegian than English in general and when speaking. Both groups also reported that they would choose to speak Norwegian more often than English when talking to a person who speaks both languages equally well. The experimental group reported reading in English more often than reading in Norwegian and the opposite is true for the overall group. Patterns of current exposure (e.g., in the last month) were similar for interacting with friends and family (more Norwegian), reading (slightly more English), television/streaming (more English) and music/media (more English). The overall group reported more self-instruction in general and especially in Norwegian. This suggests that the question may have been misinterpreted by some respondents in the overall group as the question specifies that self-instruction specifically pertains to language learning courses or apps, and it is unlikely that native speakers would do this all of the time, as indicated by the 0-10 range in responses to this question. In addition, participants were instructed that the maximum total for each row was 10 (e.g., if you speak Norwegian half of the time (= 5), you can speak English the other half of the time (= 5) or less if you also speak other languages). For several types of exposure in the overall group the total is more than 10. Finally, both groups reported more intentional and accidental use of English when speaking Norwegian than the other way around. They also reported more intentional mixing, e.g., using English words to better convey a message or describe something when their interlocutor speaks both Norwegian and English, than accidental mixing, e.g., involuntary intrusions of English when talking Norwegian. This suggests that for many of the participants, even when the language environment is mostly Norwegian, English intrudes into their L1 and is used intentionally for communicative purposes.

Proficiency ratings and questions relating to accent and dialect for the overall group are summarised in Table 4 and data from the experimental group only is found in Table 5. In both groups, average proficiency ratings are higher for Norwegian than English in all domains. Average Norwegian proficiency ranges from values corresponding to “very good” to “excellent” on the rating scale and average English proficiency ranges from “good” to “very good”. Average responses to questions about dialects, accents, and attitudes to spoken English are similar in both groups.

Summary

Overall, this paints a clear picture of L1 dominant participants with a relatively high proficiency in their L2. At the same time there is sufficient variation to look at individual differences. The overall and experimental group responses are similar enough that it was deemed appropriate to factor analyse the whole dataset to extract the factors, and to use that information when looking at the relationship between bilingual profile and behavioural data in the experimental group.

Table 2
Ratings of language exposure and mixing - all participants

Measures	L1 *M*	L1 *SD*	L1 Range	L2 *M*	L2 *SD*	L2 Range
Age milestones (years)						
Started hearing this language	0.14	0.61	0-4	6.14	2.77	0-16
Fluent speaking	4.30	2.10	1-15	12.77	3.30	4-20
Began learning to read	5.31	1.13	2-8	7.62	1.81	4-16
Fluent reading	8.06	1.85	3-20	12.13	2.52	6-20
Immersion duration (years)						
In a country	22.84	3.48	16-36	0.59	1.84	0-16
In a family	22.75	3.57	16-35	1.03	4.34	0-32
Contribution to language learning[b]						
Interacting with friends	7.85	2.58	0-10	5.47	3.00	0-10
Interacting with family	9.23	1.71	0-10	2.57	3.11	0-10
Reading	7.31	2.41	0-10	7.33	2.28	0-10
School and education	7.92	2.30	0-10	7.82	2.31	0-10
Self-instruction	1.36	2.51	0-10	2.23	2.92	0-10
TV/streaming	4.59	2.98	0-10	7.92	1.88	1-10
Music/media	3.46	3.01	0-10	7.03	2.41	0-10
Total exposure - Relative time (%)						
Exposure (general)	62.40	15.38	10-90	34.48	14.13	9-90
Exposure (speaking)	82.32	15.76	10-100	16.27	14.13	0-90
Exposure (reading)	51.93	27.15	0-97	46.64	27.08	0-100
Choose speaking this language	83.30	22.80	0-100	15.35	21.34	0-100
Exposure in the last month[c]						
Interacting with friends	8.22	1.99	1-10	2.16	2.00	0-10
Interacting with family	9.02	2.21	0-10	0.65	1.58	0-10
Reading	4.96	2.73	0-10	5.57	2.58	0-10
Self-instruction	3.37	2.43	0-10	1.71	3.06	0-10
Watching TV / streaming	2.55	2.27	0-10	7.19	2.08	1-10
Listening (music/media)	8.22	1.99	1-10	7.39	2.46	0-10
Language mixing[c]						
Accidental L2 intrusion into L1	2.88	2.46	0-10	-	-	-
Accidental L1 intrusion into L2	-	-	-	1.62	1.87	0-10
Intentional mixing L2 in L1	3.42	2.47	0-10	-	-	-
Intentional mixing L1 in L2	-	-	-	1.85	2.08	0-10

[b] Scale provided for rating contribution to language learning: 0 = not a contributor, 5 = moderate contributor and 10 = most important contributor.
[c] Scale provided for ratings of exposure in the last month and frequency of language mixing: 0 = never, 5 = half of the time and 10 = all of the time

Table 3
Ratings of language exposure and mixing - experimental group

Measures	L1 *M*	L1 *SD*	L1 Range	L2 *M*	L2 *SD*	L2 Range
Age milestones (years)						
Started hearing this language	0	0	0-0	6.03	2.74	0-15
Fluent speaking	4.05	1.85	1-10	12.08	3.07	6-19
Began learning to read	5.03	1.09	2-7	7.13	1.49	4-12
Fluent reading	8.08	2.20	5-20	11.53	2.47	8-18
Immersion duration (years)						
In a country	23.27	4.15	16-35	0.83	2.26	0-16
In a family	22.89	4.27	17-35	1.03	4.40	0-32
Contribution to language learning[b]						
Interacting with friends	7.30	2.48	2-10	5.88	2.69	0-10
Interacting with family	9.07	1.77	1-10	2.43	3.03	0-10
Reading	7.23	2.30	2-10	8.03	1.97	3-10
School and education	7.50	2.25	0-10	7.62	2.16	1-10
Self-instruction	0.30	1.09	0-6	1.75	2.89	0-10
TV/streaming	4.10	2.84	0-10	7.85	1.74	3-10
Music/media	2.77	2.68	0-10	6.55	2.56	0-10
Total exposure - Relative time (%)						
Exposure (general)	58.03	16.02	10-90	37.00	13.85	10-70
Exposure (speaking)	77.92	19.35	10-100	19.67	18.20	0-90
Exposure (reading)	43.45	25.53	3-95	55.48	25.80	4-97
Choose speaking this language	68.87	26.65	0-100	27.89	24.90	0-100
Exposure in the last month[c]						
Interacting with friends	7.65	1.85	2-10	2.20	1.79	0-7
Interacting with family	9.18	2.06	0-10	0.43	1.16	0-6
Reading	4.02	2.22	0-8	5.85	2.24	2-10
Self-instruction	0.03	0.26	0-2	0.57	1.85	0-10
Watching TV / streaming	2.67	1.69	0-6	6.93	1.93	1-10
Listening (music/media)	2.20	1.86	0-7	7.20	2.40	0-10
Language mixing[c]						
Accidental L2 intrusion into L1	2.75	2.34	0-10	-	-	-
Accidental L1 intrusion into L2	-	-	-	1.23	1.65	0-10
Intentional mixing L2 in L1	3.72	2.30	0-10	-	-	-
Intentional mixing L1 in L2	-	-	-	1.53	1.75	0-10

[b] Scale provided for rating contribution to language learning: 0 = not a contributor, 5 = moderate contributor and 10 = most important contributor.
[c] Scale provided for ratings of exposure in the last month and frequency of language mixing: 0 = never, 5 = half of the time and 10 = all of the time

Table 4
Ratings of proficiency and dialect and accent questions - all participants

Measures	L1 *M*	L1 *SD*	L1 Range	L2 *M*	L2 *SD*	L2 Range
Self-reported proficiency[a]						
Speaking (general fluency)	9.72	0.57	8-10	7.72	1.46	3-10
Pronunciation (accent)	9.62	0.71	6-10	7.04	1.60	0-10
Reading	9.42	1.10	4-10	8.07	1.50	3-10
Writing	8.86	1.22	3-10	7.31	1.56	2-10
Grammar	8.54	1.29	5-10	6.95	1.64	2-10
Vocabulary	8.47	1.20	4-10	6.86	1.57	2-10
Spelling	8.61	1.30	4-10	6.89	1.68	2-10
Dialect and accent						
L1 dialect importance[d]	6.62	3.09	0-10	-	-	-
L1 Exposure to other dialects (years)	6.32	6.59	0-30	-	-	-
L1 Modify dialect[d]	4.01	2.99	0-10	-	-	-
L1 Regional rating dialect[d]	6.98	2.63	0-10	-	-	-
L2 self-rated degree of accent[e]	-	-	-	3.38	2.10	0-10
L2 non-native perceived by others [c]	-	-	-	5.69	2.91	0-10
L2 accent importance[d]	-	-	-	7.01	2.59	0-10
L2 accent effort[f]	-	-	-	5.52	3.04	0-10
Ability to imitate accents[g]	-	-	-	4.86	2.42	0-10
Attitudes to spoken English[h]						
It is important to speak grammatically correct English	-	-	-	7.89	2.12	0-10
I pay attention to how others pronounce words and sounds	-	-	-	7.65	2.45	0-10
I want to improve my pronunciation of English	-	-	-	8.09	2.46	0-10
I would like to pronounce English like a native speaker	-	-	-	8.68	2.36	0-10
Pronunciation is not important because it does not affect communication	-	-	-	3.18	2.86	0-10

[a] Scale provided for rating proficiency: 0 = none; 1 = very low; 2 = low; 3 = fair; 4 = slightly less than adequate; 5 = adequate; 6 = slightly more than adequate; 7 = good; 8 = very good; 9 = excellent; 10 = perfect.
[c] Scale provided: 0 = never, 5 = half of the time and 10 = all of the time
[d] Scale provided: 0 = not at all, 5 = moderately, 10 = extremely
[e] Scale provided: 0 = none, 1 = almost none, 2 = very light, 3 = light, 4 = some, 5 = moderate, 6 = considerable, 7 = heavy, 8 = very heavy, 9 = extremely heavy, 10 = pervasive.
[f] Scale provided: 0 = no effort at all, 5 = moderate effort, 10 = constant effort
[g] Scale provided: 0 = extremely poor, 5 = moderate, 10 = extremely good
[h] Scale provided: 0 = very strongly disagree, 10 = very strongly agree

Table 5
Ratings of proficiency and dialect and accent questions - experimental group

Measures	L1 *M*	L1 *SD*	L1 Range	L2 *M*	L2 *SD*	L2 Range
Self-reported proficiency[a]						
Speaking (general fluency)	9.61	0.67	8-10	8.00	1.30	3-10
Pronunciation (accent)	9.57	0.79	6-10	7.20	1.40	3-10
Reading	9.43	1.00	5-10	8.42	1.42	5-10
Writing	8.75	1.13	5-10	7.62	1.43	4-10
Grammar	8.48	1.26	5-10	7.28	1.50	3-10
Vocabulary	8.48	1.02	6-10	7.30	1.42	4-10
Spelling	8.23	1.38	5-10	7.00	1.67	4-10
Dialect and accent						
L1 dialect importance[d]	5.92	3.32	0-10	-	-	-
L1 Exposure to other dialects (years)	7.12	7.69	0-30	-	-	-
L1 Modify dialect[d]	4.10	3.09	0-10	-	-	-
L1 Regional rating dialect[d]	6.62	2.77	0-10	-	-	-
L2 self-rated degree of accent[e]	-	-	-	3.20	1.77	0-8
L2 non-native perceived by others[c]	-	-	-	5.53	2.81	0-10
L2 accent importance[d]	-	-	-	7.33	2.30	0-10
L2 accent effort[f]	-	-	-	5.62	2.69	0-10
Ability to imitate accents[g]	-	-	-	4.87	2.35	0-10
Attitudes to spoken English[h]						
It is important to speak grammatically correct English	-	-	-	8.35	1.73	2-10
I pay attention to how others pronounce words and sounds	-	-	-	8.08	2.33	2-10
I want to improve my pronunciation of English	-	-	-	8.85	2.02	2-10
I would like to pronounce English like a native speaker	-	-	-	8.73	2.25	0-10
Pronunciation is not important because it does not affect communication	-	-	-	3.10	2.90	0-10

[a] Scale provided for rating proficiency: 0 = none; 1 = very low; 2 = low; 3 = fair; 4 = slightly less than adequate; 5 = adequate; 6 = slightly more than adequate; 7 = good; 8 = very good; 9 = excellent; 10 = perfect.
[c] Scale provided: 0 = never, 5 = half of the time and 10 = all of the time
[d] Scale provided: 0 = not at all, 5 = moderately, 10 = extremely
[e] Scale provided: 0 = none, 1 = almost none, 2 = very light, 3 = light, 4 = some, 5 = moderate, 6 = considerable, 7 = heavy, 8 = very heavy, 9 = extremely heavy, 10 = pervasive.
[f] Scale provided: 0 = no effort at all, 5 = moderate effort, 10 = constant effort
[g] Scale provided: 0 = extremely poor, 5 = moderate, 10 = extremely good
[h] Scale provided: 0 = very strongly disagree, 10 = very strongly agree

4.4.2 Questionnaire - factor analysis

Data processing and the analysis were conducted in R (R Core Team, 2022), using the packages GPArotation (version 2023.3.1; Bernaards & Jennrich, 2005), psych (Revelle, 2023) and nFactors (Raîche & Magis, 2022). Throughout this thesis the terms *factor analysis* and *factors* are used to refer to the variable reduction techniques and outcomes described in this chapter, even though principal components analysis (PCA) was performed.

Data handling

The data from all respondents (N = 182) were checked for variables with little variance and very high or very low correlations between variables. Some variables were removed before looking at correlations. First, parts of the responses to the question asking about what cultures the participants identify with were removed. Participants generally listed Norwegian first and rated this with a high degree of identification. Most did not list a second culture they identify with, and amongst those who did there was a lot of variation, hence this could not be quantified in a meaningful way. Second, variables showing little variation were removed, these were the age when they started hearing Norwegian, current self-instruction in Norwegian, and self-ratings of Norwegian pronunciation. Finally, questions about time spent in different language environments were removed. There was little variation in the responses and there were concerns about the reliability of this measure, as feedback from participants indicated that it was difficult to understand what these questions were asking.

A correlation matrix was produced for the 71 remaining variables. Only variables with a correlation of r = .30 or greater with at least one other variable were included in the analysis. The following five variables were excluded for not correlating with any other variable: Participant age, degree of identification with Norwegian culture, rating of education's contribution to learning English, amount of time exposed to other dialects, and degree of L1 dialect modification when communicating with speaker of a different dialect. Variables correlating above .79 were also removed. Four English and Norwegian pairs of variables assessing relative amount of language exposure from different sources were highly correlated, which is not surprising as most participants divide their time between Norwegian and English. The variables in question were general exposure (-.90), exposure reading (-.96), exposure speaking (-.96), and, if given a choice, how often they would *choose speaking* the language in question (-.96). Here the Norwegian variables were removed as differences in L2 exposure are considered more relevant to the study. Of the English proficiency ratings, speaking and pronunciation (.80), and grammar and spelling (.79), were highly correlated. Pronunciation and spelling were kept in the analysis as the questions were asking about a more specific domain and the variables in question concern lower-level processing which is considered to be more relevant to the current study. The reduced dataset containing 61 variables served as input for the analyses.

The Kaiser-Meyer-Olkin (KMO; Kaiser, 1970) measure of sampling adequacy was .70 and Bartlett's test of sphericity (Bartlett, 1950) was significant ($\chi^2(1830) = 5754.72$, $p < .001$), in line with recommendations (Williams et al., 2010, p. 5). A new correlation matrix was produced for the 61 variables. Even after removing variables correlating above .79 the determinant remained very small (2.309042e-16) indicating multicollinearity, however this may not be a problem for PCA (Field, 2013, p. 686). Individual KMO values were not considered for this analysis initially, but they should have been. At a later stage the analysis was repeated[5] after removing variables with KMO values below the .5 limit (Field, 2013, p. 706). This slightly improved the determinant, but it was still much lower than the recommended threshold of 0.00001 (Field, 2013). The analysis presented below pertains to the dataset with 61 variables.

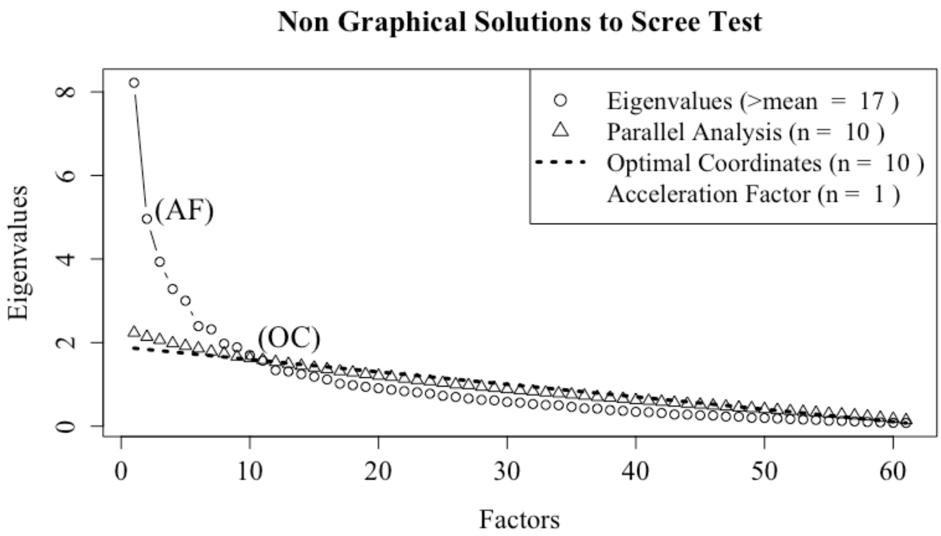
Analysis

First a PCA with 61 unrotated components was performed and the results were used to determine the number of factors to retain in the analysis. The *Kaiser criterion* (Kaiser, 1960), retaining factors with eigenvalues greater than one, and the *scree test* (Cattell, 1966), plotting eigenvalues in descending order and examining where the values level off, are probably the two most frequently used procedures for determining the number of factors (e.g., Raîche et al., 2013). *Parallel analysis* (Horn, 1965) has been found to be more reliable compared to the most frequently used measures (Zwick & Velicer, 1986), but there is a risk of overfactoring associated with all of these approaches (Fabrigar & Wegener, 2012). Figure 5 shows the scree plot and the number of factors to retain with the Kaiser

[5] Details of the additional analyses with the reduced dataset can be found in Appendix F and this is also discussed later in this chapter.

criterion and parallel analysis. One criticism of the scree test is the subjectivity associated with deciding where the plot levels off, and Figure 5 also shows two non-graphical solutions to the scree test, optimal coordinates and an acceleration factor (see Raîche et al., 2013).

Figure 5
Scree plot with the number of factors to retain using eigenvalues, parallel analysis, optimal coordinates and an acceleration factor



Note. OC = optimal coordinates, AF = acceleration factor.

In the current study the sample size is relatively small (N = 182). Field (2013) writes that the Kaiser criterion can be considered reliable with a sample size larger than 250 and average communalities ≥ .6, and that using the screeplot is recommended with a sample size greater than 200. The inflection point of the screeplot indicates retaining 12 factors[6], and the Kaiser criterion indicates retaining 17 factors. As neither the sample size nor average communalities (M = 0.55) meet these criteria, and there is a risk of overfactoring, the final decision was to extract 10 factors, as indicated by parallel analysis and optimal coordinates. First an unrotated analysis with 10 factors was performed. Considering the residuals, this seems to be an appropriate number of factors to retain. The residuals appear to be normal, indicated by the histogram of the residuals printed in Appendix E, Figure

[6] Inflection point found using uik() from the *inflection* package (Christopoulos, 2022).

E1. The root means squared residual was 0.05 and the proportion of residuals exceeding 0.05 was 0.28 ($N = 516$). The off-diagonal fit was 0.92.

The next analysis with 10 factors and oblique oblimin rotation showed no correlations between components at or above .32 (M = .08; range = .01 - .24) and following recommendations an orthogonal varimax rotation was performed (Tabachnick & Fidell, 2007). The correlations between components are printed in Appendix E, Table E1. Variable groupings and loadings for the 10 factors found in the final analysis with orthogonal rotation are listed in Table 6. The reliability of the factors was assessed with Cronbach's alpha coefficient (α; Cronbach, 1951). A general recommendation is that the value should be .7 or higher (Finch, 2020). The proportion of variance, cumulative variance, Cronbach's alpha and squared multiple correlation (SMC) for each factor are listed below the factor loadings.

The factors account for 55% of the variance in the dataset. This is not very high, but the factor structure does seem to capture variance and simplify the questionnaire data in a meaningful way. The cut-off loading value for retaining variables was set at the more inclusive .3, following Field (2013, p. 692). It has also been shown that the stability of factors depends both on the size of the loadings and the sample size (Guadagnoli & Velicer, 1988). The authors argue that loadings over .60 are reliable with a sample size of 150 or more, but with four or more loadings over .60 the results can be considered reliable with any sample size. However, with smaller loadings (.40 or under) a larger sample is recommended (at least 300-400).

In the following the results from the factor analysis and the naming of the factor constructs will be discussed in relation to these values. As described above, the results are presented in Table 6.

Table 6
Factor names and factor loadings

1. General L2 proficiency	**Loading values**	**2. L2 accent and interest**	**Loading values**	**3. L2 exposure and mixing**	**Loading values**	**4. General L1 proficiency**	**Loading values**
L2 proficiency vocabulary	0.80	L2 accent important	0.76	L2 exposure reading	0.68	L1 proficiency writing	0.86
L2 proficiency pronunciation	0.74	Effort improving L2 accent	0.73	Time reading L2 (%)	0.67	L1 proficiency spelling	0.82
L2 proficiency writing	0.73	Want to improve L2 pronunciation	0.72	General L2 exposure	0.64	L1 proficiency grammar	0.78
L2 proficiency spelling	0.70	Pay attention to pronunciation	0.68	L2 exposure friends	0.62	L1 proficiency reading	0.69
L2 proficiency reading	0.65	L2 important speaking grammatically correct	0.64	Time speaking L2 (%)	0.54	L1 proficiency vocabulary	0.69
Friend L2 learning	0.39	Want native-like L2 accent	0.61	Choose speaking L2	0.44	L2 proficiency spelling	0.41
Time speaking L2 (%)	0.39	Ability to imitate accents	0.54	Accidental L2 intrusion into L1	0.43		
Time reading L2 (%)	0.36	Pronunciation not important	-0.53	Intentional L2 mixing into L1	0.43		
L2 important speaking grammatically correct	0.35			Reading L2 learning	0.39		
Reading L2 learning	0.34			L2 Exposure TV	0.36		
Family L2 learning	0.33			L1 exposure friends	-0.30		
Choose speaking L2	0.32			L1 exposure reading	-0.42		
L2 accent perceived non-native	-0.49			L1 regional dialect	-0.47		
Accidental L1 intrusion into L2	-0.57						
Heavy accent	-0.65						
% variance	**0.09**		**0.07**		**0.07**		**0.06**
cumulative variance	**0.09**		**0.16**		**0.22**		**0.29**
Cronbach's α	**0.64**		**0.83**		**0.71**		**0.83**
SMC	**0.90**		**0.84**		**0.86**		**0.84**

5. Language development	Loading values	6. L1 Informal exposure	Loading values	7. Informal learning	Loading values	8. Informal exposure	Loading values
L2 age fluent speaking	0.77	L1 exposure TV	0.74	TV L2 learning	0.66	L2 exposure music	0.77
L2 age fluent reading	0.70	L1 exposure music	0.74	Music L2 learning	0.63	L1 exposure family	0.63
L2 age start reading	0.62	Music L1 learning	0.49	Reading L1 learning	0.54	L1 exposure friends	0.63
L2 age start hearing language	0.61	L1 exposure reading	0.40	Music L1 learning	0.51	L2 exposure TV	0.60
L1 age fluent speaking	0.53	Self-instruction L1 learning	0.39	TV L1 learning	0.49	Music L2 learning	0.39
L1 age start reading	0.51	TV L1 learning	0.36	School L1 learning	0.46	L1 exposure reading	0.33
L1 age fluent reading	0.51	L1 exposure family	0.36	Friend L1 learning	0.39		
		L2 exposure family	-0.31	Family L1 learning	0.33		
				L2 accent perceived as non-native	0.33		
% variance	**0.06**		**0.05**		**0.05**		**0.04**
cumulative variance	**0.34**		**0.39**		**0.44**		**0.48**
Cronbach's α	**0.75**		**0.75**		**0.69**		**0.67**
SMC	**0.78**		**0.78**		**0.74**		**0.70**

9. Self-instruction	Loading values	10. L2 interaction and mixing	Loading values
Self-instruction L2 learning	0.79	Intentional L1 mixing into L2	0.72
L2 exposure self-instruction	0.69	L2 exposure family	0.57
Self-instruction L1 learning	0.60	Intentional L2 mixing into L1	0.47
Family L1 learning	-0.41	Accidental L1 intrusion into L2	0.46
		Family L2 learning	0.45
		L2 accent perceived as non-native	-0.34
% variance	**0.04**		**0.04**
cumulative variance	**0.52**		**0.55**
Cronbach's α	**0.64**		**0.48**
SMC	**0.58**		**0.55**

Note. Factors are presented in descending order of variance explained and variables are sorted by loading value.

Factor descriptions
The factors were evaluated looking at the direction of the variable loadings (either positive or negative), the size of the loadings and factor reliability. Positive loading values indicate that higher variable values are positively correlated with the factor, while negative values indicate that there is an opposite relationship between the factor and the variable. The factors were named after evaluating the variables that load onto the factor.

The first factor accounts for the largest portion of variance (9%). The 5 variables with the largest positive loading values (all >.60) reflect self-reported L2 proficiency. There are also several positive loadings in the .40 to .30 range. Five relate to time spent using the L2 and more learning of L2. Specifically, time spent reading and speaking L2, and learning L2 from friends, family and reading. Choosing to speak the L2 when speaking to a person equally proficient in L1 and L2, and finding it important to speak grammatically correct L2, also load positively onto this factor. The largest negative loading (.65) is found for self-reported degree of non-native accent. Frequency of accidental L1 intrusions into L2 and frequency of being identified as a non-native speaker also load negatively onto this factor. The overall alpha for this factor was a bit low, however it has several substantial loadings related to L2 language proficiency. This factor was therefore given the name *General L2 proficiency*.

The two next factors both account for the second largest portion of variation (7% each) and were named *L2 accent and interest* and *L2 exposure and mixing*. The former contains six positive loadings above .60. Four of these variables relate to L2 accent, those are: finding L2 accent important, effort improving L2 accent, wanting to improve L2 accent, and wanting a native-like L2 accent. The desire to speak grammatically correct English also loads positively onto the factor. Two variables that may index phonological awareness/ability in general also load positively onto this factor (paying attention to pronunciation and ability to imitate accents). There is a negative loading for not finding pronunciation important. Since the factor contains different variables associated with L2 accent it was given the name *L2 accent and interest*. Interestingly, the variable specifically asking participants to rate their proficiency in L2 pronunciation loads onto the *General L2 proficiency* factor and did not group with the other variables in this domain.

The third factor contains positive loadings of seven variables relating to L2 exposure and negative loadings of two variables relating to L1 exposure. Four variables relating to L2 exposure have loadings above .60. While most variables are related to more L2 exposure, there are two positive loadings for both intentional and accidental intrusions of L2 into the L1. Therefore, this factor was named *L2 exposure and mixing*.

Factors 4 and 5 account for 6% of the variance each and were named *General L1 proficiency* and *Language development.* Five variables relating to L1 proficiency with a loading value above .60, as well as L2 spelling proficiency (.41) all load positively onto the factor that was named *General L1 proficiency.* The next factor contains positive loadings for variables relating to when participants reached specific language development milestones. The four L2 variables all have loading values above .60, while the three L1 variables (one was excluded prior to analysis) have loading values above .50. Since both L1 and L2 variables showed relatively high loadings, this factor was named *Language development*.

The next two factors account for 5% of the variance each and were named *L1 informal exposure* and *Informal learning.* The first factor contains two positive loadings above .60 for L1 exposure from music and television. L1 exposure through reading and L1 learning through music have positive loadings above .40. In the lower range there are positive loadings for learning L1 through self-instruction and TV, as well as a positive loading for L1 exposure from family. The one negative loading was for L2 exposure through family. The next factor contains a mix of variables mostly related to language learning. Two variables reflecting L2 learning from TV and music show high positive loadings (>.60), while four variables reflecting L1 learning show positive loadings above .40. Finally, L1 learning from friends and family, and L2 accent being perceived as non-native have positive loadings in the .40 to .30 range. This factor is less clear than the ones that have been discussed so far and was given the language non-specific name *Informal learning*.

The last three factors each explain 4% of the variance. Factor 8 was given the name *Informal exposure* as it contains four positive loadings above .60 for L2 exposure through music and television and L1 exposure through family and friends. There are also two positive loadings in the .40 to .30 range for L2 learning through music and L1 exposure from reading. The next factor was named *Self-instruction.* Self-instruction for learning L1 and L2 and current exposure to L2 through self-instruction have loadings at or above .60, and there is a negative loading for learning L1 through family. The final factor was given the name *L2 interaction and mixing.* This factor had the lowest reliability. The only loading above .60 was the positive loading of intentional language mixing of L1 into the L2. Intentional language mixing of L2 into the L1, accidental L1 intrusions into L2 and both learning and exposure to L2 through family all had positive loadings above .40. Finally, there was a negative loading for L2 accent perceived as non-native.

Factor scores were calculated using the default regression approach in the psych package, which is parallel with the approach used for factor analysis. Individual factor scores were extracted for the participants who completed all experiments in the current study and serve as measures of bilingual profile and individual difference predictors in this thesis. As mentioned earlier, individual KMO values were not considered in the analysis presented here. After completing the rest of this project, variables with KMO values below .5 were removed and new analyses were conducted using the reduced dataset. The results from two new analyses with an orthogonal varimax rotation and additional information can be found in Appendix F. The first analysis had 10 factors to mirror the original analysis (Appendix F, Table F2). With the new dataset, parallel analysis and optimal coordinates indicated that 9 factors should be retained. Therefore, a second analysis with 9 factors is also reported (Appendix F, Table F3).

The five first factors are similar across analyses in terms of variable structure and loadings. This allowed for a comparison of factor scores extracted from the old analysis and the new analyses, and they were found to be highly correlated (see Appendix F, Table F1). The last factors were not very similar across analyses. There were no clear improvements in reliability or interpretability of the last factors in the new models compared to the original model, therefore it was deemed appropriate to retain the factor scores from the original analysis.

Summary and comparison to original LEAP-Q

The main goals of the factor analysis were to reduce the complexity of the dataset and to see how questionnaire variables group together. The 10 factors found in the current analysis accounted for 55% of the variance in the dataset. Variable groupings generally seemed to reflect different aspects of bilingual profile, but the five first factors were generally the most stable and reliable. Several of the questions that were added to the LEAP-Q in the current study group together in the factors *L2 accent and interest, L2 exposure and mixing* and *L2 interaction and mixing*, suggesting that the addition of these variables successfully expand the questionnaire to include the intended domains of accent and phonology and language switching experience. The original LEAP-Q was developed and tested on bilingual participants with more diverse language backgrounds and in a different language environment compared to participants in the current study. The results from the current factor analysis do seem to reflect a different "type" of bilingualism. This suggests that the questionnaire successfully captures critical differences between different populations. The findings from the current study and those reported for the original LEAP-Q are compared in more detail below.

The factor analysis reported in Study 1 in the original LEAP-Q paper (Marian et al., 2007), with bilinguals who spoke English combined with various languages, obtained 8 factors, while the current study found 10 factors. In both studies, variables assessing L1 and L2 proficiency (*competence* in the original paper) grouped together. Variables which grouped into two factors called *Media-based learning* and *L1 maintenance* in the original study, generally overlap with three factors in the current study: *L1 Informal exposure, Informal learning*, and *Informal exposure*. This seems to reflect the different language environments of the participants in each study, i.e., immersion or living in an L2 environment in the original study, compared to living in an L1 environment in the current study.

In the current study, the factor *Language development* contains positive variable loadings for both L1 and L2 language learning milestones, and the *Self-instruction* factor also applies to both languages. The original study saw late learning of L2 group together with less time spent in the L2 country and more self-instruction (use of language tapes). Marian et al. (2007) additionally found factors reflecting late L2 immersion, balanced immersion, and non-native status, neither of which apply to the participants in the current experiment. This suggests that the factors found in the original study reflect immersion in an L2 country, while the most important factors for learning in the current study seem to reflect individual differences in language learning that affect both the L1 and the L2.

In the second study reported in the original paper (with bilinguals speaking English and Spanish, one group with L1 Spanish and another with L1 English), the 8 factors found reflected slightly different constructs. The first factor suggested that relative competence in the two languages accounted for the most variation, reflecting the different dominance profiles of the participants, while in the current study participants are more uniformly L1 dominant. Similar to Study 1 in Marian et al. (2007), there were factors reflecting immersion and language dominance and media-based learning, and finally a new variable associated with less identification with L2 culture and L2 acquisition starting at a later age, which again is less likely to apply to the current study where participants live in an L1-dominant environment and L2 acquisition generally started at the same time for all participants.

4.4.3 Language tests - results, evaluation of tests, and comparisons with self-ratings

All tests were created (or substantially changed from the original) for this experiment, apart from the English versions of the elision and serial nonword recognition task. The goal was to create tests measuring specific language skills and to avoid ceiling performance. This section will start by evaluating the experiments themselves, before comparing self-ratings to test scores. All data processing and statistical analysis were conducted in R (R Core Team, 2022) and the alpha was set at .05. Figures were made with the packages ggplot2 (version 3.4.2; Wickham, 2016) and gridExtra (Auguie, 2017).

Participants

The 60 participants (referred to as experimental group in Section 4.4.1) in this study were all native speakers of Norwegian with English as their strongest L2 (44 female, 16 male). The mean age was 24.10 years (*SD* = 4.25) and they had 16.35 years of education on average (*SD* = 2.33).

Data handling

Prior to the analysis the data were examined and analysed to determine which analysis methods were appropriate. Data from three of the language tests (L1 spelling test and elision in both languages) were skewed and failed to meet assumptions required for a Pearson correlation. Some data points that could be considered outliers were identified in the elision tasks through visual inspection of boxplots. All test values accurately reflect each participant's score on the given test and represent "… legitimate cases sampled from the correct population" and therefore using a transformation to keep the observations is recommended (Osborne & Overbay, 2004, pp. 2–3). Several transformations that may reduce skewness were attempted (e.g., Baayen, 2008, p. 92; Mangiafico, 2016, pp. 703–721), however two of the datasets remained skewed. The positively skewed data from the Norwegian version of the spelling test were successfully normalised using a square root transformation, but the negatively skewed results from both versions of the elision task either remained negatively skewed or became positively skewed depending on the transformation. The data were therefore kept non-transformed and cross-language comparisons of test scores (e.g., comparing scores from the L1 spelling test and L2 spelling test) are reported using both Pearson and Kendall correlation measures, since Kendall makes no assumptions about normality.

Cross-language comparisons of language tests

Test accuracy (mean, standard deviation, and range) and the results from paired correlations (Pearson and Kendall) are all reported in Table 7. The means are relatively similar in Norwegian and English for all tests apart from the spelling test. The relationship between Norwegian orthography and phonology is more transparent (or more shallow) than for English (Seymour et al., 2003, p. 146) and this might explain the larger cross-linguistic difference between the spelling means compared to the other tests. The significant between-language correlations suggest that the tests measure similar competencies in both languages (assuming these are related) and that if a person scores highly on a test in one language, it is likely that the score in the other language will be high as well.

Table 7
Paired language test results, descriptive statistics

	Test accuracy (100% = 1.0)						**Correlation coefficients**			
	Norwegian			**English**			**Pearson**		**Kendall**	
TEST	*M*	*SD*	range	*M*	*SD*	range	$r(58)$	p	r_τ	p
Spelling	0.64	0.48	0.40-1.00	0.46	0.50	0.05-0.95	.59	< .001	.40	< .001
Vocabulary	0.40	0.49	0.13-0.70	0.42	0.49	0.13-0.80	.37	< .01	.23	< .05
Elision	0.89	0.31	0.69-1.00	0.90	0.29	0.69-1.00	.62	< .001	.53	< .001
SNWR	0.66	0.47	0.41-0.89	0.64	0.48	0.41-0.89	.26	< .05	.20	< .05

The range of scores suggests that the tests were generally successful in producing a spread of values and avoiding ceiling performance. However, for the Norwegian spelling test and elision in both languages some participants were able to respond correctly on all trials. The mean accuracy and range on the Norwegian spelling test suggest that only a few participants scored at ceiling level, but for the elision task mean accuracy is high for both languages and the range is smaller compared to the other tests. This might suggest that the elision task is too easy in its current form and not an ideal measure for investigating individual differences in this population.

While the test scores are similar in general it is important to note that the results are not directly comparable. For instance, obtaining a score of 60% correct on both the Norwegian and English vocabulary test does not mean that a person's vocabulary proficiency is the same in both languages. As discussed previously (Section 4.2.1), no two tests in two different languages measure the exact same thing, and that is not the claim here either. However, the tests generally seem suitable for investigating the relationship between measured and self-rated language performance, as well as individual differences.

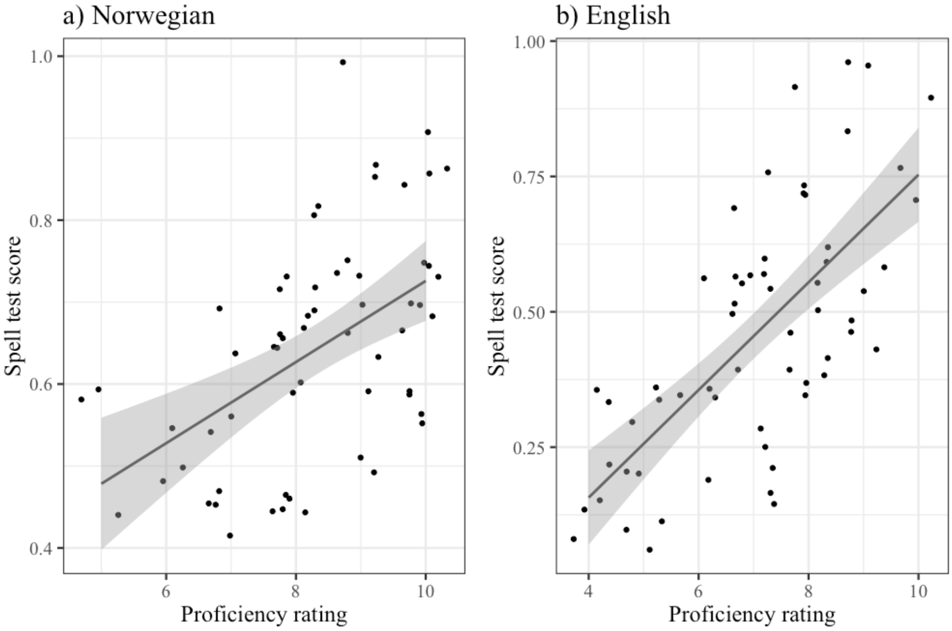
Comparing self-reported and objective measures

Several studies report correlations between self-reported proficiency and measured proficiency. Kendall's correlation coefficient[7] was calculated to investigate the relationship between the self-reported proficiency and test results obtained in this study. Figure 6 shows the relationship between self-ratings of spelling proficiency and test scores in Norwegian (a) and English (b). Visually there seems to be a relationship between ratings and scores in both languages, and that the correlation is stronger for English. There is a statistically significant correlation between ratings and test scores both in English ($r_\tau = .57, p < .001$) and Norwegian ($r_\tau = .41, p < .001$). Figure 7 shows the relationship between self-ratings of vocabulary proficiency and test scores in Norwegian (a) and English (b). Here the correlation between English ratings and scores is significant ($r_\tau = .42, p < .001$), but not for Norwegian ($r_\tau = .15, p = .151$).

[7] Only Kendall's tau is reported here. In addition to the problems with meeting assumptions, there are also different measurement levels for test scores (0.00 -1.00) and self-ratings (0 - 10).

Figure 6

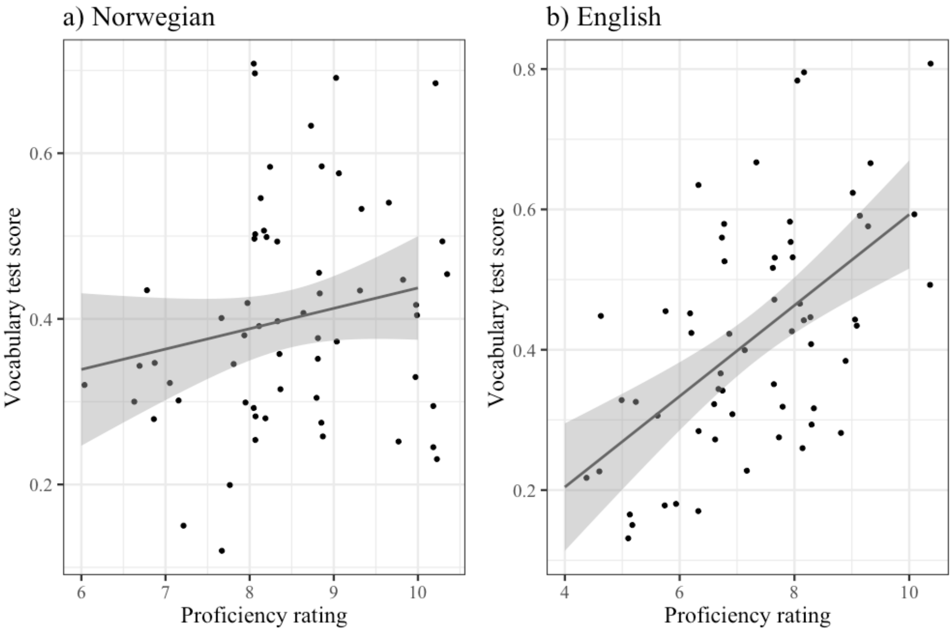
Scatter plot with regression line for self-reported spelling proficiency and spelling test scores in Norwegian and English



Note. The shaded area shows the standard deviation.

Figure 7

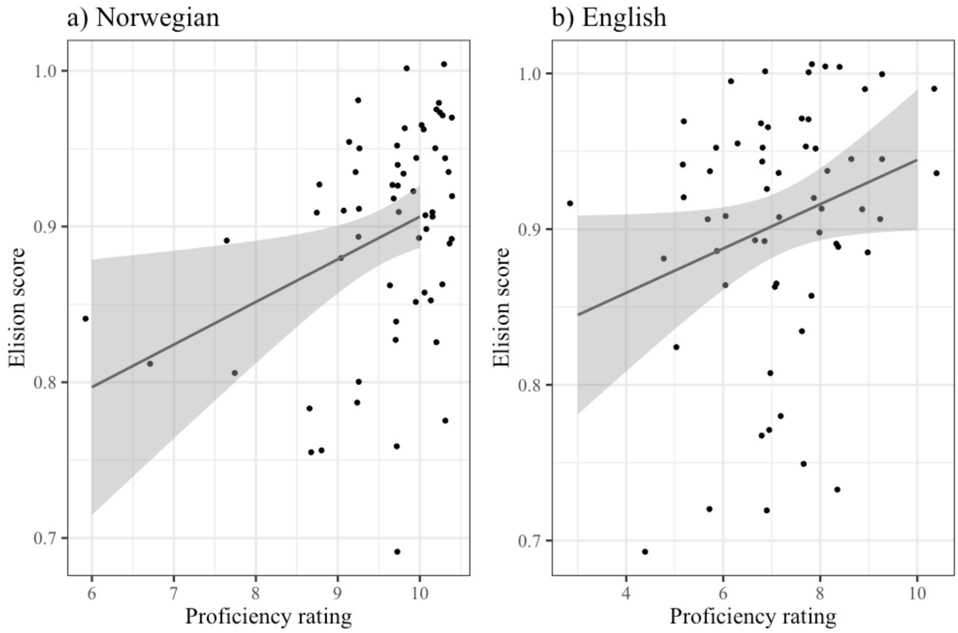
Scatter plot with regression line for self-reported vocabulary proficiency and vocabulary test scores in Norwegian and English



Note. The shaded area shows the standard deviation.

Figure 8

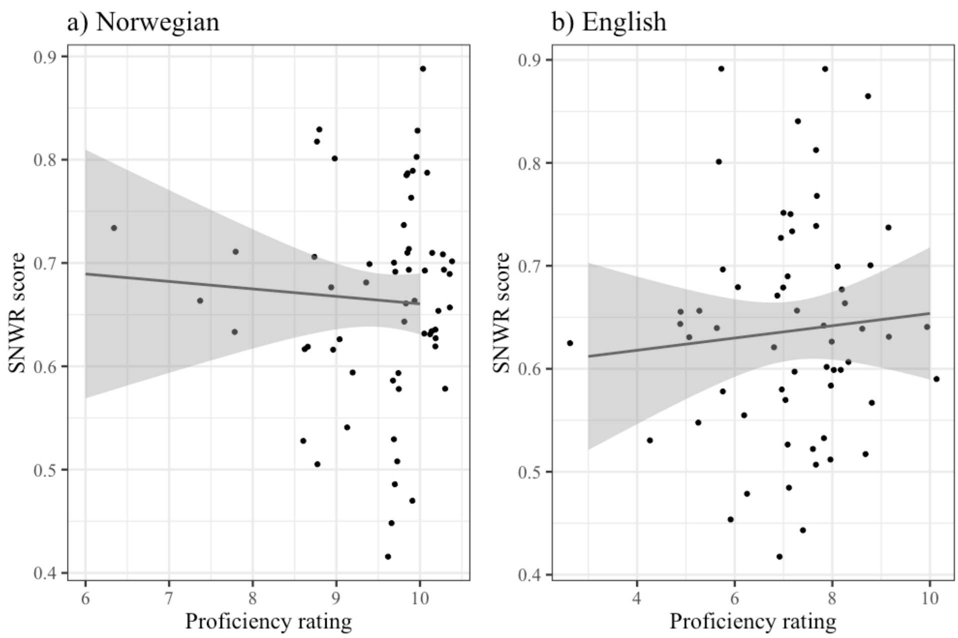
Scatter plot with regression line for self-reported proficiency pronouncing each language and elision test scores in Norwegian and English



Note. The shaded area shows the standard deviation.

Figure 9

Scatter plot with regression line for self-reported proficiency pronouncing each language and serial nonword recognition (SNWR) test scores in Norwegian and English



Note. The shaded area shows the standard deviation.

In addition to the spelling and vocabulary tests participants also completed two tests related to phonological processing: elision (Figure 8) and SNWR (Figure 9). Participants were not explicitly asked to rate their "phonological proficiency". This domain is more difficult to define than spelling or vocabulary, and it is unclear to what extent people are aware of their abilities in this domain. However, there seems to be some overlap between speech perception and production (see e.g., Section 2.5). Self-rated pronunciation proficiency in Norwegian and English was therefore correlated with elision and SNWR scores in the respective languages. SNWR scores did not correlate with the self-ratings in either language, but there was a significant correlation between elision scores and self-ratings of pronunciation in both Norwegian ($r_\tau = .25, p < .05$) and English ($r_\tau = .20, p < .05$).

Summary

The results partially support previous findings of a relationship between self-rated and tested proficiency. Generally, the relationship is stronger between English (L2) self-rated and tested proficiency than for Norwegian, similar to what Marian et al. found (2007, p. 960). Possible explanations for this difference include characteristics of the tests themselves and that the participants might be more self-aware, or have received more feedback, on their L2 proficiency compared to their L1 proficiency. For the phonological measures, elision and SNWR, only elision test scores correlated with self-rated pronunciation proficiency.

4.4.4 Comparing language test results and factors extracted from the questionnaire

This section reports the results from multiple linear regressions with the test scores against the factors extracted from the questionnaire. Multiple regressions were run in R (R Core Team, 2022) for each language test against all 10 factors extracted from the questionnaire. The alpha was set at .05. The participants are the same as in Section 4.4.3, i.e., the 60 participants who completed all experiments.

Data handling

The test scores and LEAP-Q variables were mean-centred prior to the analysis. The variance inflection factor (VIF) was below the recommended threshold of 3 for all covariates (Zuur et al., 2010, p. 9), indicating that there are no problems with multicollinearity. VIF values for covariates are reported in the result tables below. The assumptions for regression were checked following Winter (2013).

Visual inspection of quantile-quantile plots and residual plots of the residuals did not indicate that the assumptions normality and homogeneity of variance were violated for the vocabulary, spelling, or serial nonword recognition tasks in both languages and the English elision task. Residuals from the Norwegian elision task were skewed. Log and square root transformations did not improve skewness or change the strength of the correlations. Results are therefore reported from the analysis with untransformed data.

Spelling

Table 8 shows the results from two multiple regressions, the Norwegian and the English spelling test with the factors extracted from the LEAP-Q. A significant regression equation was found for both the English ($F(10,49) = 4.31$, $p < .001$) and Norwegian ($F(10,49) = 2.83$, $p = .007$) analysis. Adjusted R^2 was .360 and .237, respectively. Ordered by strength of the relationship, the factors *General L2 proficiency*, *L2 exposure and mixing*, *General L1 proficiency* and *L2 accent and interest* showed significant positive association with English spelling scores. For Norwegian, *General L1 proficiency* and *Informal learning* were significantly associated with test scores.

In general, the spelling test results are associated with the factors one might expect, provided that the self-ratings and the factor groupings are reliable. Both language proficiency and language exposure factors are positively associated with spelling test results. English scores are also linked to *L1 proficiency*, which supports the previous findings of correlations between Norwegian and English spelling test scores, as well as between self-reported proficiency and test scores in this domain. The *Informal learning* factor associated with Norwegian test scores contains ratings of how different types of exposure contribute to L1 and L2 learning. It contains both music and television for both languages, and school and reading for Norwegian.

Table 8
Significant results from multiple regressions of Norwegian and English spelling test scores against the factors extracted from the questionnaire

	Norwegian spelling				English spelling				VIF
Factor name	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	
General L2 proficiency					.1085	.0261	4.16	< .001	1.11
L2 accent and interest					.0687	.0311	2.21	.032	1.38
L2 exposure and mixing					.0916	.0288	3.18	.003	1.33
General L1 proficiency	.0750	.0194	3.87	< .001	.0729	.0300	2.43	.019	1.47
Informal learning	.0581	.0186	3.13	.003					1.14

Vocabulary

The results from regression with Norwegian and English vocabulary tests against the factors extracted from the LEAP-Q are found in Table 9. A significant regression equation was found for English ($F(10,49) = 6.17$, $p < .001$) but not Norwegian ($F(10,49) = 0.93$, $p = .513$). Adjusted R^2 was .469 and -.012, respectively.

Ordered by magnitude of influence, the factors *General L2 proficiency* and *L2 exposure and mixing* were associated positively with English test scores, similar to the spelling test scores. There is a significant negative relationship to both *L1 Informal exposure* and *Self-instruction*, as well as a marginally negative relationship with *General L1 proficiency*. Less exposure to the L1 generally means more exposure to the L2 and the association to English vocabulary scores is therefore not surprising. The *Self-instruction* factor contains both English and Norwegian ratings, but it is not significantly related to Norwegian vocabulary scores, and in fact no factors were significant for Norwegian. In Section 4.4.3 we saw that self-ratings and Norwegian vocabulary test scores were not correlated either. Why self-instruction is negatively associated with English vocabulary scores is unclear. One possible explanation could be that while most participants reported little to no self-instruction, the ones who did might be less proficient language users in general and therefore engage in self-instruction more frequently to enhance their language abilities.

Table 9
Significant results from multiple regressions of Norwegian and English vocabulary test scores against the factors extracted from the questionnaire

	Norwegian vocabulary				English vocabulary				VIF
Factor name	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	
General L2 proficiency					.0799	.0170	4.71	< .001	1.11
L2 exposure and mixing					.0774	.0188	4.13	< .001	1.33
General L1 proficiency					-0.0359	.0195	-1.84	.072	1.47
L1 Informal exposure					-0.5992	.0259	-2.31	.025	1.92
Self-instruction					-0.0606	.0252	-2.40	.020	2.01

Note. Marginal correlation in grey.

Elision

The results from regression with Norwegian and English elision scores against the factors extracted from the LEAP-Q are found in Table 10. A significant regression equation was found for the English ($F(10,49) = 3.4$, $p = .002$) and the Norwegian ($F(10,49) = 2.55$, $p = .015$) analysis. Adjusted R^2 was .289 and .208, respectively. Ordered by magnitude of influence, English elision scores were positively associated with *General L2 proficiency*, *L2 exposure and mixing*, and *L2 interaction and mixing*.

For English elision, as for spelling and vocabulary, higher *General L2 proficiency* and *L2 exposure and mixing* relate to higher test scores. The factor *L2 interaction and mixing* seems to reflect more L2 exposure growing up, and perhaps more balanced bilinguals, as it contains both L1 substitutions into L2 and L2 into L1, as well as more L2 interaction with family compared to *L2 exposure and mixing.* The latter only contains L2 substitutions when speaking in their L1 and L2 interaction is mostly with friends. This suggests that there is a relationship to general proficiency, but in addition this may reflect the association between early language exposure, language development and phonological processes. It should be noted, however, that this was the least reliable factor. The link between elision scores and early language development becomes more apparent looking at the Norwegian results, which is not surprising as participants generally were more exposed to Norwegian at a young age.

Table 10
Results from multiple regressions of Norwegian and English elision test scores against the factors extracted from the questionnaire

	Norwegian elision				English elision				VIF
Factor name	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	
General L2 proficiency					.0332	.0095	3.51	< .001	1.11
L2 accent and interest	.0267	.0106	2.51	.015					1.38
L2 exposure and mixing	.0205	.0098	2.09	.042	.0283	.0105	2.71	.009	1.33
Language development	-.0234	.0112	-2.08	.043					1.25
Self-instruction	-.0321	.0132	-2.42	.019					2.01
L2 interaction and mixing					.0206	.0098	2.09	.042	1.11

Higher Norwegian elision scores relate to *L2 accent and interest* and *L2 exposure and mixing*, while there is a negative relationship with *Self-instruction* and *Language development.* The negative association with language development suggests that participants who reported learning and mastering speaking and reading at a younger age scored higher on the Norwegian elision task than participants who reported reaching these milestones later. Better performance on phonological (working) memory tasks is associated with benefits in early language development in both L1 and L2 (e.g., Dufva & Voeten, 1999; Gathercole et al., 1992), and it has been suggested that phonological memory is a more influential predictor of early language development than language performance at later stages (Gathercole et al., 1992). In addition, if more *Self-instruction* is related to less proficiency, as suggested in the discussion of the vocabulary results above, this also supports a relationship to early exposure and proficiency. This might also explain the fact that higher values of the *L2 accent and interest* factor are associated with higher scores on the Norwegian elision task, but not the English elision task.

Serial nonword recognition (SNWR)

Neither the English ($F(10,49) = 1.52$, $p = .160$) nor the Norwegian ($F(10,49) = 0.91$, $p = .532$) regression equations were significant and adjusted R^2 was .081 and -.016, respectively. Overall, little variation was explained by the predictors. The only significant predictor for either language was a negative relationship between English SNWR scores and the *Language development* factor (Estimate = -0.047, $SE = 0.019$, $t = -2.52$, $p = .015$).

The negative association between language development and English SNWR scores implies a role of phonological proficiency in early stages of language development. However, compared to the elision results the SNWR scores show a weaker association to the factors in general. This supports the notion that these two tests are measuring different aspects of phonological processing and may reflect the difficulty of obtaining self-reported measures of phonological proficiency and awareness compared to measures in other language domains. As mentioned, these two tests also place different additional demands on the participant. The elision task involves articulatory demands, working memory and phonological awareness, while the SNWR test mainly involves phonological short-term memory.

Summary of all regressions

Table 11 provides an overview of the results from all regressions. Factors are presented in descending order of variance explained in the factor analysis (Section 4.4.2). In general, factors are more associated with L2 performance than L1 performance and the weakest relationship between factors and test scores overall is observed for the SNWR task.

Table 11
Overview of all significant regressions (t-values) with test scores against factors obtained from the questionnaire

Factor name	Language and test							
	L1 Spelling	L2 Spelling	L1 Vocabulary	L2 Vocabulary	L1 Elision	L2 Elision	L1 SNWR	L2 SNWR
General L2 proficiency		4.16		4.53		3.51		
L2 accent and interest		2.21			2.51			
L2 exposure and mixing		3.18		4.13	2.09	2.71		
General L1 proficiency	3.87	2.43		-1.84				
Language development					-2.08			-2.52
L1 Informal exposure				-2.31				
Informal learning	3.13							
Informal exposure								
Self-instruction				-2.40	-2.42			
L2 interaction and mixing						2.09		

Note. Marginal correlation in grey.

The first factor, *General L2 proficiency*, is associated with performance on three of the L2 tests: spelling, vocabulary, and elision. *L2 accent and interest* is associated with L2 spelling scores and L1 elision scores which could be expected as this factor contains variables relating to phonology and accent, most of which were added to the questionnaire to obtain measures for these aspects of language use. However, it was not a significant predictor for L2 elision scores nor the SNWR scores.

L2 exposure and mixing parallels *General L2 proficiency* in predicting L2 test results, but the relationship is slightly weaker. In addition, it is associated with L1 elision scores. *General L1 proficiency* is not as successful in predicting L1 test scores as the *General L2 proficiency* variable is at predicting L2 performance.

General L1 proficiency predicts both L1 and L2 spelling scores, but the effect is stronger for L1. None of the other L1 test are significantly related to the L1 proficiency factor. Finally, there was a negative relationship between general L1 proficiency and L2 vocabulary scores.

The remaining factors are generally associated with fewer test scores and the reason for their relationship is sometimes less apparent, but as discussed previously these might be related to amount and type of exposure to each language. The *Language development* factor negatively predicts SNWR scores in L2 and elision scores in L1. *L1 Informal exposure* reflects more exposure to Norwegian though television, music and reading, both currently and growing up, and therefore possibly less input of L2 vocabulary. This factor is negatively associated with L2 vocabulary scores. *Informal learning* predicts L1 spelling scores and no other scores in either language. *Informal exposure* is a general factor consisting of the same amount of L2 and L1 variables which does not predict performance on any of the tests. *Self-instruction* is negatively associated with performance on L2 vocabulary and L1 elision. As discussed previously, higher amounts of self-instruction may reflect less language proficiency. Finally, *L2 interaction and mixing*, is positively associated with L2 elision and may reflect higher amounts of L2 exposure at earlier stages of language development.

4.5 Discussion

This chapter has reported an investigation into the relationship between self-reported and tested variables associated with the domains of phonology and accent and other aspects of bilingual profile. Two research questions were posed at the beginning of this chapter:

1. What is the relationship between self-ratings in the domains of phonology and accent and self-ratings of other areas of bilingual profile?
2. Is there a relationship between self-ratings in the domains of phonology and accent and objective measures of phonological processing?

The participants completed an extensive questionnaire and four language tests, both in their L1 and L2. These measures provide a detailed assessment of bilingual language profile. This chapter evaluated the extension of a bilingual questionnaire, augmented to include more questions related to the domains of phonology and accent, language switching and more detailed assessments of language proficiency. Specifically, the interest lies in whether these questions add more dimensions to the description of bilingual profiles or if variation along these dimensions is already accounted for by general language proficiency and other aspects of bilingual profile. This is formalised in the research question: what is the relationship between self-ratings in the domains of phonology and accent and self-ratings of other areas of bilingual profile? The factor *L2 accent and interest* contained seven variables relating to accent and phonology and is one of two factors accounting for the second largest proportion of variance (7%) in the dataset. This emerges as a clear accent factor that is different from other measures of bilingual language profile. In addition, some variables, perhaps unsurprisingly, also relate to general proficiency, learning and language use. Four variables relating to phonology and accent load onto the factor *General L2 proficiency,* and the variable *L2 accent perceived as non-native* loaded positively onto the factor *Informal learning* and negatively onto the factor *L2 interaction and mixing.* This suggests that the questions added to the questionnaire contribute to a new accent factor, but that this construct is not completely separate from general L2 proficiency and factors relating to language exposure.

The relationships between domain specific self-ratings of proficiency and language test scores were also assessed to investigate the link between self-reported and tested proficiency in the current study. Self-ratings of spelling and pronunciation proficiency were correlated with spelling and elision scores, respectively, in both languages. Self-ratings of vocabulary proficiency and vocabulary scores only correlated for English and self-ratings of pronunciation proficiency and SNWR scores did not correlate in any of the languages.

To further analyse the measures of proficiency, separate multiple regressions were conducted for each objective language test and the factors extracted from the questionnaire. The second research question, asking whether there is a relationship between self-ratings in the domains of phonology and accent and objective measures of phonological processing, cannot be answered definitely here. The results suggest that self-ratings correlate with elision scores, as mentioned above. That is, there is a correlation between elision scores and self-rated pronunciation proficiency in both languages, as well as positive relationship between L1 elision and the *L2 accent and interest factor*. On the other hand, there was no significant relationship between L2 elision and the *L2 accent and interest* factor, and the

SNWR scores were not related to the self-ratings of pronunciation proficiency, nor the *L2 accent and interest* factor. As discussed previously, these tests involve different additional demands for the participant, the elision task involving articulatory demands, working memory and phonological awareness, while the SNWR test, especially the longest test items, places high demands on phonological short-term memory. The current results could therefore reflect that these two tests measure different aspects of phonological processing.

Phonological memory has been implicated as an important factor in early language development (Dufva & Voeten, 1999; Gathercole et al., 1992). In the current study, the only significant result with the SNWR task is a negative correlation between L2 SNWR scores and the *Language development* factor, suggesting that participants who reported reaching language milestones at an earlier age, in both L2 and L1, scored higher on the L2 SNWR task compared to participants who reached those milestones later. This could indicate that differences in phonological short-term memory are most influential in early language development, but only to a small extent do they reflect current differences in this population of proficient L2 speakers. For adult learners of an L2, better performance on an SNWR task has been found to correlate with better L2 spoken fluency (O'Brien et al., 2007) and more native-like perception of speech sounds (Cerviño-Povedano & Mora, 2015). It is therefore possible that SNWR measures might correlate with the more fine-grained phonetic measures employed in Chapter 6 of this thesis.

It is also worth considering what self-ratings in the domains of phonology and accent entail. As mentioned, eliciting self-ratings in the phonological domain is not straightforward and the questionnaire focused on attitudes to accent and "phonology-adjacent" questions, such as effort and interest in improving L2 accent, ability to imitate accents, and whether the participants pay attention to accent in speech. The first measure considered was self-ratings of pronunciation in L1 and L2, both of which correlated with elision in the respective languages, but not SNWR. As the elision task involves articulation and the ability to manipulate speech sounds, the link to proficiency in pronunciation is more direct compared to the SNWR which requires participants to remember and compare two strings of nonword syllables. Even though the SNWR test involves higher memory demands for the participant compared to many other tests, it is often chosen over other tests specifically because it does not rely on articulatory processes and as such is a more pure measure of phonological processing (e.g., Cerviño-Povedano & Mora, 2015). The term *domains of phonology and accent* is used as a broad label in this thesis, intended to cover both self-reported variables related to accent and phonology as well as the associated language tests and articulation measures employed in Chapter 6. The results above suggest that the questionnaire variables meant to address the domains of phonology and accent were more successful at capturing aspects relating to pronunciation and accent, than phonology.

The first research question addressed in this chapter asked about the relationship between self-ratings in the domains of phonology and accent compared to self-ratings in other domains. While a clear accent factor emerged, there was some overlap with other domains. This was also the case for the elision task results which were not only associated with the *L2 accent and interest* factor and self-ratings of pronunciation proficiency, but also general language proficiency and language experience, perhaps reflecting its role in language development. That is, in both languages, higher scores on the elision task were associated with higher loadings on the *L2 exposure and mixing* factor, the factors *General L2 proficiency* and *L2 interaction and mixing* predicted higher L2 elision scores and L1 elision scores were related to factors suggesting a benefit in language development (i.e., the negative relationships with the factors *Language development* and *Self-instruction*). The other factors were associated with test scores to some extent, apart from *Informal exposure*, and in general the factors that explained the most variance in the factor analysis were associated with more of the test scores. Overall, factors were more associated with L2 performance than L1 performance and the weakest relationship between factors and test scores is observed for the SNWR task.

The paired objective language tests employed in this thesis were evaluated by looking at correlations between test scores in each language and the degree to which each test produced a spread of values while avoiding ceiling performance. All comparisons were significant, and generally the tests produced spread and avoided ceiling performance. However, some participants had an accuracy rate of 100% for the Norwegian spelling test and the elision task in both languages. While only few participants reached ceiling level on the spelling test, the elision task might have been too simple and in the future the results could possibly be refined by looking at reaction times. Another potential area of improvement is the

Norwegian version of the vocabulary task, which showed no correlations with other measures.

The Norwegian vocabulary test appeared to be the most challenging of the objective tests with a maximum score of 70% correct, while the rest of the tests in both languages had a maximum score of 80% correct (English vocabulary) or more (see Table 7 for full overview). As this test was developed for this thesis, future work could seek to develop a set of validated and more comparable tests for Norwegian-English bilinguals. A future study might benefit from adjusting the difficulty level by replacing some of the more challenging Norwegian test words with more frequent words. As older adults tend to score higher on vocabulary tests (Verhaeghen, 2003), it would also be of interest to see how older adults perform on the same test.

In general, the expectation was that participants would rate their native language proficiency highly, however, self-ratings of proficiency in Norwegian range from 5 (*adequate*) to 10 (*perfect*) (see Table 4 in Section 4.4.1 for full scale specification). Ratings of vocabulary proficiency range from 6 (*slightly more than adequate*) to 10. The vocabulary test was completed before self-ratings of proficiency. It is therefore possible that some participants who scored highly compared to others on the Norwegian vocabulary test still rated their proficiency as relatively low since they felt that many words were unfamiliar. Section 4.2.1 discussed some factors that may affect self-ratings, such as proficiency level (Trofimovich et al., 2016) and anxiety associated with speaking in the L2 (MacIntyre et al., 1997), which could have influenced the current results. When comparing the self-ratings of the 5 participants with the highest and lowest accuracy on the vocabulary tests, the Norwegian self-ratings are relatively similar in the two groups (lowest: $M = 8.4$, Range = 7 - 10; highest: $M = 8.8$, Range = 8 - 10). In English there is a larger difference between the groups with the lowest ($M = 5.4$, Range = 5 - 6) and highest accuracy on the test ($M = 8.4$, Range = 7 - 10). It is not possible to tell whether a slightly easier Norwegian vocabulary test would have led certain participants to rate their vocabulary proficiency higher, but this is something to keep in mind for future studies, both when developing tests and when deciding on response scales in questionnaires.

4.5.1 Conclusion

This chapter has focused on ways of assessing language background and proficiency in bilinguals. The literature review shows that there are several methods available and a trend towards more detailed assessment in recent years, for instance by combining (standardised) test scores and self-ratings of proficiency, language background, and language experience. In the current study participants completed four language tests in each language and an extensive questionnaire, augmented to include self-ratings in the domains of phonology and accent. The objective language tests that were developed for this experiment only correlated with self-ratings for some domains, and both language tests and self-reported measures seem to be more reliable for L2.

The questionnaire data were factor analysed and the added variables generally group together and contribute to a clear accent factor. The results therefore indicate that the inclusion of self-ratings in the domains of phonology and accent contribute to a richer assessment of bilingual profile overall, however there is also some overlap between domains. This chapter also evaluated the relationships between two objective measures of phonological processing, elision and SNWR, and two self-rated measures, self-ratings of pronunciation proficiency and the *L2 accent and interest* factor. There was a positive correlation between elision scores and pronunciation proficiency in both languages. Furthermore, higher L1 elision scores were associated with higher values of the *L2 accent and interest* factor. On the other hand, SNWR scores did not correlate with self-ratings, nor the factor, for either language. While the elision results suggest that there is a link between performance and self-ratings in the domains of phonology and accent, this does not extend to phonological short-term memory, as measured by the SNWR task, in the current population.

The factor analysis yielded ten factors that contribute to the description of bilingual profile for the participants in this project. The five first factors were generally the most reliable in terms of alpha and the size of the factor loadings. Factor scores from two of these factors, *General L2 proficiency* and *L2 accent and interest* will be used to investigate effects of individual differences in bilingual word production and language control in the following chapter. Then Chapter 6 of this thesis will investigate the relationship between individual differences in articulation, the bilingual profile measures addressed in the current chapter, and word production and language control.

5 Individual differences in bilingual word production and language control

5.1 Introduction

The word production process, as discussed in Chapter 2, has been described as serial activation of distinct components that generate the speech output. Different models make different assumptions about the exact nature of the stages in this process and how information flows from one stage to the next, but there is general consensus that there are at least two key stages involved: conceptualising and formulation (e.g., Levelt, 1999). The first referring to the message or content of what the speaker wants to say, and the second to creating a linguistic structure. Research suggests that bilingual word production is affected by non-selective activation of the bilingual two languages (e.g., Green, 1998; Hoshino & Kroll, 2008; Meuter & Allport, 1999), as well as differences in bilingual profile, such as language exposure and proficiency (e.g., Bonfieni et al., 2019; Costa & Santesteban, 2004). Non-selective language activation in turn, means that there is a need for language control during bilingual language production (e.g., Declerck & Philipp, 2015). Finally, the prevalence of cognate effects in both perception and production, even between language with different scripts (e.g., Gollan et al., 1997; Hoshino & Kroll, 2008), suggests that there is non-selective or cross-linguistic activation of phonology.

The effect of individual differences in bilingual profile on word production and language control is the focus of the current chapter. As mentioned above, bilingual word production is affected by language proficiency and cognate facilitation effects suggest a role for phonology. In the current study I report the results from a simple picture naming task (completed in both L1 and L2) and a picture naming task with language switching, both of which include a cognate manipulation (experiments addressed in this chapter are framed in Figure 10). Two factors from the factor analysis in Chapter 4 were added as predictors in the analysis of latencies and errors from the naming tasks to investigate the role individual differences. These are the *General L2 proficiency* factor which is a more general L2 proficiency factor including several domains and the *L2 accent and interest* factor which is more related to the domains of phonology and accent. Performance on a version of the flanker task (Zhou & Krott, 2018) is included as a non-linguistic measure of selective attention and control.

Figure 10
Experiment overview Chapter 5

Day 1 - single language	Day 2 - single language	Day 3 - both languages
Picture naming	Picture naming	Switching task
Reading	Reading	Flanker task
Elision	Elision	Questionnaire
Spelling	Spelling	Experiment debrief
Serial nonword recognition	Serial nonword recogntition	
Vocabulary	Vocabulary	
Language counterbalanced - either L1 - L2 or L2 - L1		

Note. Experiments addressed in Chapter 5 are framed.

This chapter starts by reviewing empirical findings on bilingual word production and language switching. First the definition of cognates and previous studies of cognate effects in simple naming are discussed. This is followed by a review of general language switching studies and language studies that specifically have manipulated cognate status. Throughout possible mechanisms for language control are discussed. Then the method for each experiment, including the design and materials are discussed. The results from the picture naming and language switching experiments are described separately, followed by a general discussion of the results.

In addition to investigating naming and switching behaviour in a new population of bilinguals, two research questions relating to individual differences in language production are addressed in this chapter:

- How do individual differences in non-linguistic attentional control interact with cognate effects in language production and switching?
- Do aspects of bilingual profile modulate naming and switching behaviour?

In the analysis, previously reported main effects of cognate facilitation, response language, and trial type were replicated. Cognate effects were uniformly helpful in both picture naming and language switching. The results also support and extend findings of global L1 inhibition in language control. Switch costs were symmetrical, and a reversed dominance effect was observed in the switching task. Individual differences in non-linguistic attentional control interacted with cognate effects in the picture naming task, but not the switching task. There were no significant effects of the *General L2 proficiency* predictor in picture naming nor the language switching task. The *L2 accent and interest* predictor main effect was significant in the picture naming task, as well as in a significant interaction with cognate status and naming language. In the switching task both predictors were involved in marginal interactions. The limited effects observed were generally in the direction of faster production with higher values of the predictors, however considering the marginal and complex nature of the interactions, there is no strong evidence of the predictors modulating switching behaviour.

5.2 Theoretical overview

5.2.1 Cognates and cognate effects

This thesis has readily adopted the definition of cognates frequently used in a psycholinguistic context (e.g., Costa et al., 2000, p. 1285; Hoshino & Kroll, 2008, p. 503), i.e., words sharing form and meaning across languages. In philology this definition is reserved for native words sharing the same origin, and does not include loan words (Minkova, 2014). In this view, cognates may no longer share form or meaning in a transparent way, for instance the false friends Norwegian *kinn* (/çin/, meaning cheek in English) and English *chin* (/tʃɪn/). It is also important to note that for cognates, from here on referring to cognates in a psycholinguistic sense, the semantics associated with the word form might be similar across languages but not completely overlapping. For instance, the meaning could be broader in one language compared to the other (e.g., “pudding” which can describe sweet and savoury dishes in both Norwegian and British English, but in the latter also could mean dessert in general) or associated with both different and overlapping meanings (e.g., Norwegian “under” translates to both “under” and “wonder” in English). Conceptual differences between languages are not expected to be an issue in this thesis where participants are mostly naming simple objects. Finally, cross-language differences in sound inventories, phonotactics and orthography can influence the degree of similarity between cognates. Nevertheless, evidence suggests that cognates, due to their cross-linguistic similarity and representational convergence, affect bilingual speech production, as described below.

Costa et al. (2000) report two experiments showing robust cognate facilitation effects that are modulated by language dominance. In the first experiment 21 highly proficient Catalan-Spanish bilinguals and 21 Spanish monolinguals named 40 pictures in Spanish. The bilinguals reported using both languages daily, but Catalan was the dominant language. They manipulated both frequency and cognate status, as frequency should affect both monolingual and bilingual responses, while the cognate distinction should only affect bilingual naming. The stimuli were 10 low-frequency cognates, 10 high-frequency cognates, 10 low-frequency noncognates and 10 high-frequency noncognates (each repeated 16 times across 4 blocks). Similar frequency effects were obtained for both groups, but cognate facilitation was observed for the bilingual group only. In the second experiment, the effect of language dominance was assessed and 46 Catalan-Spanish bilinguals, 23 of which were Spanish dominant and 23 who were Catalan dominant, named pictures in Spanish. The procedure was the same as in the first experiment, but more pictures were added (80 pictures in total). Cognate facilitation was found in both groups, but the effect was larger for the Catalan dominant group who named pictures in their non-dominant language. The modulation of cognate facilitation by language dominance has also been found for accuracy rates on the Boston Naming task, were balanced Spanish-English bilinguals showed similar cognate facilitation in both languages, while facilitation was stronger in the nondominant language for unbalanced bilinguals (Rosselli et al., 2014).

A study with 27 Japanese-English bilinguals and 35 Spanish-English bilinguals found evidence of cognate facilitation in picture naming for bilinguals who speak languages with different orthographic scripts (Hoshino & Kroll, 2008). The stimuli consisted of 72 items, half of which were cognates, and stimuli were repeated four times. In addition, cognate type was manipulated, meaning that words could be a cognate between all three languages, in Japanese and English, or in Spanish and English. They found cognate facilitation for both groups, suggesting that phonological similarity speeds spoken word production irrespective of script type and providing further evidence for non-selective language activation.

5.2.2 Language switching and language control

Cognate facilitation in speech production supports non-selective language activation. The presence of non-selective language activation, in turn, suggests that there is a need for a control mechanism in bilingual language production. Several studies have used picture naming with cued language switching to investigate the consequences of non-selective activation and possible mechanisms for language control.

In the seminal study by Meuter and Allport (1999), language switching was studied by presenting digits (1-9) to 16 unbalanced bilinguals. The colour of a frame surrounding the digits cued which language the participants should respond in. Participants were bilinguals from different language backgrounds, but all of them spoke English as either their L1 or L2. Each participant completed approximately 2000 trials (half in L1 and half in L2). The trials were distributed in lists varying lengths (from 5 to 14 trials in each list), the number of same-language runs before a switch ranged from 1 to 13, and the number of switches within a list ranged from 0 to 4. The results show a consistent pattern of higher switch costs when switching from a weaker L2 into a stronger L1 than vice versa. These findings are in line with the predictions of the inhibitory control model (ICM) (Green, 1998) described in Section 2.4. That is, in bilingual lexical access the dominant language is generally more strongly activated, and therefore more strongly inhibited, when naming in the nondominant language than vice versa. This inhibition in turn makes it more difficult to switch back into the dominant language, resulting in asymmetrical switch costs. Post hoc they divided the participants into two groups of less balanced and more balanced bilinguals, based on the relative naming speed in L1 and L2 on nonswitch trials. The interaction between relative proficiency, trial type and language was not significant, however they still ran separate analyses for each group which suggested that the switch cost asymmetry was attenuated for the more balanced group.

These findings have been replicated and extended in five language-switching experiments reported in Costa and Santesteban (2004) and four further experiments in Costa et al. (2006). In both studies the trial structure was kept similar to Meuter and Allport (1999), but there were fewer trials (950 in total) and picture stimuli with noncognate names instead of digits, as the number of digits that are cognates will depend on the combination of languages being studied. The first experiment in Costa and Santesteban (2004) replicated the asymmetrical switch costs, in the direction of larger switch costs for L1 than L2, in two groups of unbalanced bilinguals (Spanish-Catalan and Korean-Spanish, 24 in total). In the second and third experiment, participants were two different groups of 12 Spanish-Catalan bilinguals who were highly proficient in their L2 (Catalan). The same general procedure was followed, but in Experiment 3 an additional 30 noncognate pictures were added and there were fewer repetitions of each stimulus. Here, symmetrical switch costs were found in both experiments. In Experiment 4, Spanish-Catalan bilinguals (who matched the participants in Experiments 2 and 3) switched between their L1 Spanish and their much weaker third language (L3) English, using the same materials as in Experiment 1. For these proficient bilinguals, switch costs were symmetrical even when switching from a weaker L3. In addition, they found that in Experiments 2 - 4 overall response times were shorter in L2 than L1 (reversed dominance). In the final experiment they manipulated preparation time, by presenting the language cue 500 ms or 800 ms before the stimulus, in order to investigate whether the longer L1 latencies were the result of a bias towards L2 selection. That is, if L2 naming is faster because participants by default prepare to name in L2 and thereby facilitate L2 naming, this difference should disappear when participants know the naming language in advance. More preparation time reduced switch costs, but L2 naming latencies were still faster than L1. The authors argue that the lack of asymmetric switch costs in Experiment 4, and the faster responses in L2, cannot readily be accounted for assuming an inhibitory control mechanism. They hypothesise that for highly proficient bilinguals, language control is instead achieved by a mechanism setting a language specific threshold of activation for lexical selection.

Costa et al. (2006) used the same experimental procedure as in Experiment 2 of Costa and Santesteban (2004) to compare the performance of the highly proficient Spanish-Catalan bilinguals in that study to two new groups of highly proficient bilinguals (11 Spanish-Basque bilinguals and 12 Spanish-English bilinguals). The results were similar for all groups, that is symmetric switch costs and shorter naming latencies in L2 compared to L1. In the second experiment they found symmetrical switch costs for 12 highly proficient Spanish – Catalan bilinguals who switched between their L2 (Catalan) and a weaker L3 (English). This result is discussed in the context of their previous paper (Costa & Santesteban, 2004), which proposes that highly proficient bilinguals do not rely on inhibitory processes in language control, but rather a language-specific mechanism. Given the findings of symmetrical switch costs for proficient bilinguals, even when a weaker third language was involved, they hypothesised that after there is a shift to a language-specific selection mechanism it can be applied to any of the bilingual's languages. They tested this assumption in Experiments 3 and 4. First, 12 highly proficient Spanish-Catalan bilinguals switched between their L3 (English) and their fourth language (French). In the final experiment 12 Spanish-Catalan bilinguals (similar to previous experiments) and 12 Spanish monolinguals learned 10 invented picture names and switched between their L1 (Spanish) and the invented language. In both experiments, naming was faster in the dominant language and switch costs were asymmetrical with larger switch costs for the more dominant language. The authors therefore suggest that different control mechanisms may be employed depending on the proficiency level of the bilingual speaker. That is, they assume reactive inhibition is involved in language control, but only for newly learned languages and/or weak languages, while a different language-specific mechanism for control, such as adjusting the activation threshold for selecting the L1 or a different type of inhibition, operates for more proficient bilinguals. This could account for the longer response latencies in L1 observed in the experiments with highly proficient bilinguals (Costa et al., 2006).

Mechanisms of cognitive control may be divided into two categories, proactive control and reactive control (Braver, 2012). In bilingual language control, mechanisms that are applied before lexical access can be said to exert proactive control, while mechanisms that are applied after languages are activated exert reactive control (e.g., Green, 1998; Green & Abutalebi, 2013). Mechanisms that are applied to the whole language operate globally, while local mechanisms effect smaller units, such as individual lemmas (e.g., De Groot & Christoffels, 2006). How these control mechanisms relate is not fully understood. It is feasible, however, that different mechanisms are used in language control, but that their involvement depends on the proficiency of the speaker as well as the demands of the production task (e.g., Christoffels et al., 2007; Green & Abutalebi, 2013; Timmer et al., 2019). Language control mechanisms have also been studied by manipulating preparation time, that is the time between the presentation of a (language) cue and a stimulus (Costa & Santesteban, 2004; Fink & Goldrick, 2015; Verhoef et al., 2009).

So far, the studies mentioned have analysed switch costs, i.e., the difference between switch and stay trials. Switch costs are believed to index reactive inhibition, which operates from trial to trial. Mix costs, on the other hand, are believed to index proactive control. Mix costs are obtained by comparing simple picture naming trials to stay trials in a switching task, thereby indexing the sustained differential need for control in a single language production context compared to production in language switching. The view that language control mechanisms are task dependent, and can involve both proactive and reactive control is supported by the analysis of switch and mix costs in studies that manipulate preparation time, including presenting the stimuli before the language cue (Khateb et al., 2017; F. Ma et al., 2016).

A recent study included aspects of individual differences in bilingual profile and experience as predictors in the analysis of switch and mix costs (Bonfieni et al., 2019). Participants were two groups of highly proficient bilinguals, 37 Italian-English bilinguals who acquired their L2 later than the other group, but had more daily exposure to their L2 English, and 46 Italian-Sardinian bilinguals who were more balanced. For the more balanced Italian-Sardinian group, switch costs were symmetric while switch costs were asymmetric in the Italian-English group. Interestingly, the switch cost was larger for the L2. This group was immersed in an L2 environment, and the authors speculate whether this could be related to overall higher L2 activation compared to the other group. The analysis with the predictors (combined analysis with both groups) found that higher L2 proficiency was associated with faster switch trials in both languages. For both groups mix costs were asymmetric and larger for L1. This could suggest that the different language environments to some extent work against the expected dominance effects. They found that more daily L2 exposure reduced mix costs in L1, possibly reducing the relative amount of proactive control needed to supress the L1.

To summarise, language switching studies to date show a complex pattern of results, with switch costs modulated by a number of aspects of bilingual profile. As described above, evidence that language selection in bilingual language production is non-selective suggests the involvement of a control mechanism (or mechanisms) to allow for successful language production. This notion has led to questions such as whether exercising language control affects non-linguistic cognitive functioning and to what extent there is overlap between mechanisms for general cognitive control and language control.

Several studies have compared monolingual and bilingual participants on non-linguistic cognitive tasks (e.g., Prior & Macwhinney, 2010; Woumans et al., 2015). Bilinguals have for instance been found to outperform monolinguals in the attentional network task, suggesting that non-linguistic switching is less costly for bilinguals than monolinguals (Costa et al., 2008). There are also studies showing individual differences between bilinguals. A study comparing bilinguals with strong language control abilities to bilinguals with weaker language control abilities found that the former group performed better on a selection of tests assessing executive function (Festman et al., 2010). It has therefore been suggested that there are overlapping mechanisms between bilingual language control and executive control (e.g., Declerck et al., 2017). On the other hand there is evidence suggesting that there is a difference between language control processes and general executive control, for instance the finding of differential switch costs when comparing linguistic and non-linguistic switch costs within the same population of bilinguals (Calabria et al., 2012).

The frequency with which bilinguals switch between their languages has also been shown to effect switching costs in both task and language switching, that is, smaller switch costs were found for bilinguals who report switching between their languages often, compared to monolinguals and bilinguals who switch between their languages less frequently (Prior & Gollan, 2011). In a recent review paper, de Bruin (2019) describes several sources of individual variation that have been investigated in relation to bilingual executive functioning: Age of acquisition, language proficiency, context of language acquisition, amount of language use, and how the languages are used (language switching and language context). This highlights the importance of considering individual differences when investigating bilinguals.

In the current study, the population is a relatively homogeneous group of bilinguals with a high proficiency in both languages, albeit a higher proficiency in L1. Based on the studies above we expect to find evidence of L1 inhibition in the switching task and symmetric switch costs. It remains to be seen the extent to which these measures are affected by the individual difference measures employed in this thesis. The current study was not designed to distinguish between types of control in language switching and will therefore not analyse mix costs, but reaction times and switch costs. In addition to two bilingual profile predictors, performance on a flanker task is included to assess whether observed differences in switching behaviour can be explained by general differences in non-linguistic selective attention and control. To conclude this theoretical review, I will address the limited set of switching studies that have included a cognate manipulation.

5.2.3 Cognate effects in language switching

In the studies reported so far, the trial structure has been similar, but with some variation in stimulus type (digits or noncognate pictures), preparation time, and the number of stimuli. Stimulus type has been shown to affect switch costs in that smaller switch costs have been found for digits compared to noncognates and semantically related items (Declerck et al., 2012). Interestingly, there was no significant difference between cognate items and digits, suggesting that the cross-linguistic similarity in digits (depending on the language combination) and cognates similarly affect language switching. To date only a small group of studies have investigated how cognates are selected in language switching tasks. Similar to the switching studies discussed above, different groups of bilinguals have been investigated and varying experimental designs have been used.

Declerck et al. (2012) compared how language switching is affected by stimuli type in four experimental blocks (108 trials each) with different types of stimuli: digits, noncognates, cognates, and semantically related items (9 of each). 24 unbalanced German-English bilinguals participated in the study. The results from the switching task with digits were compared individually to the three other stimulus types. They found that the results from digit naming (digits with cognate names had different degrees of phonological overlap) patterned with cognate naming (pictures of cognate words). Both of which showed smaller switch costs compared to noncognates and semantically related stimuli. The evidence for cognate and digit switch cost facilitation was interpreted as evidence for overlapping phonological representations.

In Christoffels et al. (2007) 24 German-Dutch unbalanced bilinguals, who switched between their languages daily, named 48 pictures (half with cognate names) both in single language blocks (96 trials) and in language switching blocks

(768 trials). Language cues and stimuli were presented simultaneously. They found cognate facilitation on latencies in all conditions. L1 latencies were shorter than L2 latencies in the single language blocks and this pattern was reversed in the switched blocks, mainly due to slower naming in L1, suggesting a proactive control mechanism. Finally, both switch costs and mix costs were symmetrical and language mixing affected the L1 more strongly than the L2, a finding which was especially evident in the cognate facilitation effect. The latency findings relating to language control were supported by EEG data.

Verhoef et al. (2009) manipulated the duration of the interval between language cues and stimuli (750 or 1500 ms) in a language switching task. 17 Dutch-English unbalanced bilinguals named 48 pictures (half with cognate names) across 1536 trials. There was an overall cognate facilitation effect that was stronger for the L1, and L1 latencies were longer than L2 latencies. There was no interaction between cognate status and switch costs. Switch costs were asymmetric when the preparation interval was short, but symmetric when the interval was long. Furthermore, they found that preparation times modulated latencies in general, but not for L1 stay trials. They argue that this could reflect that on L1 stay trials there is no competition for lexical selection, and therefore no need for inhibition, but that inhibition can modulate lexical selection amongst activated candidates. Accompanying EEG data were also consistent with less inhibition on L1 stay trials.

Santesteban and Costa (2016) investigated whether language switching was affected by the cognate status of both the target word and the preceding word. For half of the participants critical items were always preceded by trials of the same type (e.g., a critical cognate trial was always preceded by a cognate trial) and for the other half critical trials were always preceded by a word with the opposite cognate status. The interval between cue and target varied and 20 pictures (half cognate) were named over 950 trials in total. Two groups of bilinguals, 24 unbalanced Spanish-Catalan bilinguals and 24 highly proficient Spanish-Catalan bilinguals, participated in the study. They found main effects of cognate facilitation, asymmetric switch costs for the less balanced group and symmetric switch costs for the highly proficient group. Switch costs were not modulated by the cognate status of target or preceding word and the authors argue that while cognates facilitate naming latencies, they do not facilitate switching or alter lexical selection mechanisms.

In the studies reported so far, stimulus pictures have been repeated throughout the experiments. Broersma et al. (2016) report a language switching study where pictures were only presented once, and critical cognate stimuli and paired noncognate controls (two sets of 18 cognates and controls) were distributed amongst 177 filler trials. Language switches only occurred on filler trials. Participants were 48 highly proficient Welsh-English bilinguals (17 more dominant in Welsh, 17 more dominant in English and 14 with equal or context-dependent dominance). They found symmetric switch costs and that cognates could both facilitate and inhibit latencies, modulated by language dominance. That is, for the English dominant participants cognates inhibited production in Welsh and did not affect latencies in English, while for the Welsh dominant participants cognates facilitated production in both languages. They argue that for cognate words there could be competition for selection at lexical-semantic level, while there is facilitation at the word form level, and therefore, depending on the relative strength of activation at these levels of processing the result could be facilitation, inhibition, or no measured cognate effect if the levels of activation at the lexical-semantic level and phonological (word-form) level cancel out.

Finally, C. Li and Gollan (2018) examined cognate effects in language switching in three experiments. In the first experiment cognate status was blocked, while cognate status was mixed within blocks in the other two experiments. Stimuli and cues were presented simultaneously. Similar groups of 32 Spanish-English bilinguals (English dominant) participated in the experiments. In the first experiment, participants named cognate and noncognate words (9 of each) in two separate blocks (total 216 trials, 12 repetitions per item in each block). They found evidence of switch costs, cognate facilitation, and marginally faster latencies in the nondominant language and that cognates reduced switch costs in the dominant language. They also considered whether the repetition of items in the experiment affected latencies by analysing the first presentation of each picture (18 per participant) and obtained a similar pattern of results. Finally, they compared block order (whether participants completed the experiment in English or Spanish first), and obtained similar results, however the difference between L1 and L2 latencies was greater in the second block. The second experiment used the same general procedure and stimuli, but cognates and noncognates were mixed within each block. They found similar main effects, but switch costs were not modulated by cognate status or language. When analysing the first presentations of stimuli they found cognate facilitation of switch costs in both languages, suggesting an effect of stimulus repetition. In the final experiment the number of stimuli were reduced (four cognates and four noncognates) to assess whether the number of intervening trials between stimulus repetition affected latencies. Again, they found similar main effects (but cognate status was not significant) and evidence for cognate inhibition (larger switch costs for cognates than noncognates). On the first presentation of trials, however they found marginal cognate facilitation. These findings suggest that both stimulus repetition and experimental design can modulate cognate-switch facilitation effects.

The literature to date has found varying effects of cognate status on language switching. Some of this variability may be attributable to differences in the experimental designs used. In addition, however, like in the studies using only noncognate stimuli, individual differences in language proficiency contribute to some of the variability observed. Few studies to date have investigated individual differences in word naming and language switching. Given the important role of phonological representations in cognate processing, individual differences in the domains of phonology and accent may especially contribute to the variability of cognate effects, both in spoken word production and in language switching. In the experiments reported in this chapter, cognate status is manipulated in a spoken word production experiment (picture naming on separate days for L1 and L2) and a language switching experiment. Individual differences are assessed by including two predictors from the factor analysis in Chapter 4 (*General L2 proficiency* and *L2 accent and interest*) and performance on a version of the flanker task (Zhou & Krott, 2018).

5.3 Materials and design

In all three experiments the participants were 60 native speakers of Norwegian with English as their strongest L2 (44 female, 16 male). The mean age was 24.10 years (*SD* = 4.25) and they had 16.35 years of education on average (*SD* = 2.33). A more detailed description can be found in Section 4.4.1, where the participants in question are referred to as the experimental group. In the following the materials, design, and procedure are described in turn for the picture naming experiment, the language switching experiment, and the flanker task.

5.3.1 Picture naming

Materials

96 pictures (300 x 300 pixels) were selected from the 750 pictures in the MultiPic (Multilingual Picture) databank (Duñabeitia et al., 2018). Half with cognate and half with noncognate names (Appendix G). An additional 10 pictures with 100% naming agreement in English were used for task familiarisation.

The stimuli were selected to contain segments that may be problematic for Norwegian L2 speakers of English (when naming in L2) and L1 segments that are similar or often used to replace these problematic L2 segments (when naming in L1). The problematic L2 segments in question are /dʒ tʃ v w z s θ ɜː uː ʊ ʌ/. The voiced dental fricative is not included because there was only one occurrence in the entire MultiPic databank. The same pictures were used for both Norwegian and English picture naming. The picture names in the database have been normed for six languages, including English, but not yet for Norwegian. Therefore, pictures were assigned Norwegian names by the experimenter.

Cognates vary in degree of similarity, as discussed in Section 5.2.1. Before selecting the stimulus materials, variation in terms of orthography and phonology was coded. The semantics associated with the word, both between languages and between speakers, may also vary, but this is more difficult to control. All cognates in the final stimulus set refer to relatively simple concepts and are likely to have a high correspondence between languages.

In addition to cognate status and cognate similarity, all 750 items in the databank were coded for frequency, orthographic and phonological word length, number of syllables, and phonological neighbourhood density. Visual complexity ratings and English naming agreement percentages for the pictures were provided by the databank. Materials were selected seeking a balanced distribution of each of these factors across languages and cognate status. Tables of means are provided in Appendix H, along with within and between group comparisons (Table H1 and Table H2, respectively). The cognate and noncognate sets within each language did not differ significantly on any of these variables (all *p*s $< .05$). However, between-language comparisons showed a significant difference ($p = .042$) between the number of syllables in Norwegian ($M = 1.85$, $SD = 0.80$) and English ($M =$

1.54, *SD* = 0.68) cognates as well as log frequency values for both cognates and noncognates.

Norwegian word form frequencies were obtained from NoWaC (Guevara, 2010). American English and British English word form frequencies were obtained from SUBTLEX-US (Brysbaert & New, 2009) and SUBTLEX-UK (van Heuven et al., 2014), respectively. Phonological neighbourhood density values for the English words were obtained from The Irvine Phonotactic Online Dictionary (Vaden et al., 2009). Norwegian phonological neighbourhood density values were calculated manually following Vitevitch and Luce (1999, p. 381) where a neighbour is defined as "any transcription that could be converted to the transcription of the stimulus word by a one phoneme substitution, deletion, or addition in any position".

To compare frequency measures across languages, Norwegian frequencies were normalised to Zipf-scale values, using the formula provided by van Heuven et al. (2014, p. 1179). As mentioned, there are no significant within-language differences in word frequency across cognate status. A significant difference was found between English (both types) and Norwegian frequencies. In general, Norwegian frequencies are lower than the English frequencies. This overall trend could reflect actual differences in usage. However, the source material for the frequency counts could also be an issue. The American and British English frequencies are based on subtitles (Brysbaert & New, 2009; van Heuven et al., 2014), while the Norwegian Bokmål frequencies are based on documents downloaded from the .no domain (Guevara, 2010). The semantic content of subtitles and documents published on the internet may differ in general and therefore confound the between-language comparison of word frequencies. This is a potential confound, however, if these values do reflect higher L2 frequencies compared to L1, this works in the opposite direction of language dominance, and critically, the within language manipulations of cognate status are not affected by this. It has also been pointed out that meaningful cross-language comparisons of frequency, especially for less documented languages, may not always be possible with currently available resources (Bonfieni et al., 2019).

In other picture naming studies word onset segments are often carefully controlled. Working from a database of 750 pictures there were some aspects that were not perfectly matched between conditions. After the analysis of the picture naming task was completed, the final model was used as a starting point to assess the potential influence of onset differences (voicing and manner of articulation) between languages. The models were evaluated through backward model comparisons and Benjamini-Hochberg significance (Benjamini & Hochberg, 1995) is reported to control the false discovery rate for multiple comparisons (full models and output in Appendix I, voicing in Tables I1 and manner of articulation in Table I2). Onset voicing and onset manner of articulation did not interact with the between item variables.

Design

Four pseudo-randomised lists were created for each language with each of the 96 items appearing four times in total. First, cognate and noncognate items were sorted separately by descending SUBTLEX-UK frequencies. Alternating items were assigned to separate lists, resulting in two lists with 24 cognates and 24 noncognates of similar frequencies. Each list was pseudorandomised 4 times. Additional criteria were applied so that there were:

> (1) no more than four subsequent stimuli with the same cognate status; (2) no stimuli of the same semantic category, or semantically related ones, followed each other; (3) no stimuli names with the same phonetic onset followed each other; (4) repetition of a picture was separated by at least four intervening trials. (Zheng et al., 2018, pp. 141–142)

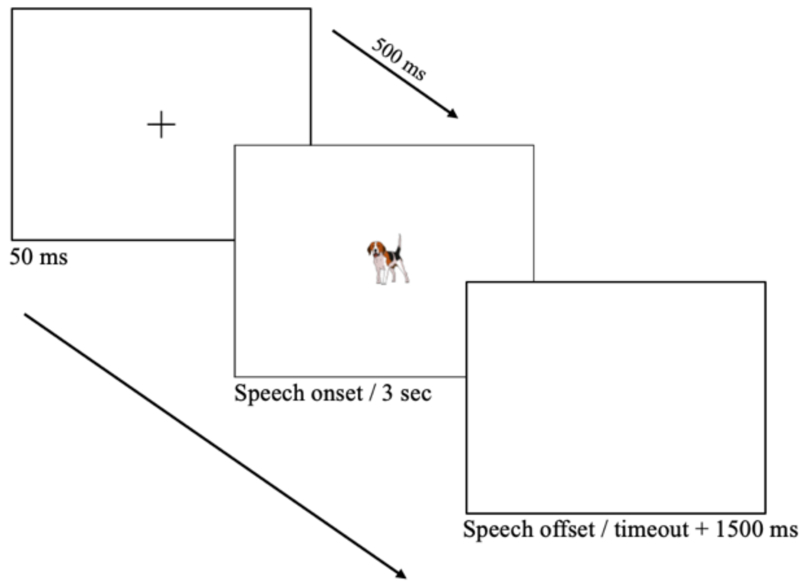
Finally, the eight pseudorandomised lists were concatenated and rotated using a Latin square design, creating four experimental lists for each language. Each list contained 384 items and was divided into 16 blocks. The same materials were used for L1 and L2 naming, but the list order was different for each language.

Procedure

Participants named pictures in L1 Norwegian and L2 English on separate days (the order was counterbalanced across participants) and there was never more than two days between sessions. The participants were familiarised with the pictures prior to the experiment proper. The 96 experimental items, and an additional 10 practice items, were presented on a screen with the correct name printed underneath in font size 48. Participants were asked to pay attention to both the picture and the name. The presentation of the items followed the same structure as the experimental trials (described below). They were familiarised with the English and Norwegian names separately, preceding naming in the respective language. Participants were told to indicate whether any of the words or pictures were odd or unfamiliar in the breaks between the presentation blocks.

After familiarisation, there was a short practice block with 10 pictures before the experiment began. The experiment consisted of 16 blocks with 24 pictures. There was a pause between each block. The trial structure is visualised in Figure 11. Each trial began with the presentation of a 50 ms beep and fixation cross placed at the centre of the screen. After 500 ms the picture to be named was presented. A voice key recorded speech onset and speech offset. The picture disappeared from the screen at speech onset. The response timeout was 3000 ms after picture onset. The next trial began 1500 ms after speech offset (or timeout). Participants were instructed to respond as quickly and accurately as possible.

Figure 11
Picture naming trial structure



5.3.2 Language switching

Materials

48 pictures (24 cognate and 24 noncognate) were selected from the Picture Naming Task materials (see Appendix G). The stimuli were controlled for the same factors as the picture naming task stimuli. There were no significant differences between cognate and noncognate items in the within-language comparisons. There was a significant difference between unstressed phonological neighbourhood density in American English and Norwegian phonological neighbourhood density and as before there was a significant cross-linguistic difference in frequency for both cognates and noncognates. Tables of means are provided in Appendix H, along with within and between group comparisons in Table H3 and Table H4, respectively.

Design

Twelve pseudorandomised lists were created with each picture appearing once per experimental condition. Trials could be either switch or stay and the naming language could be either L1 or L2, resulting in four experimental conditions. Cognate and noncognate words appeared in alternate blocks giving eight blocks in total. Each picture was named once in each condition, and once per experimental block. At the end of the session, participants were given a debrief questionnaire to check whether this manipulation was noticed by the participants. One participant reported noticing that the experiment was blocked by cognate status in the 7th block and the rest did not appear to notice at any point during the experiment.

There was an equal number of switch and non-switch trials. Pseudorandomisation followed constraints from Zheng et al. (2018), as in the picture naming task (see Section 5.3.1). The structure of the experimental list was created based on 1-3 trials naming in the same language before a switch. Eight separate randomised lists were created for L1 and L2 runs in order to create the blocks. These were interleaved resulting in a pseudorandomised experimental list with eight blocks, four starting with L1, four starting with L2. With this set-up, participants might be able to learn that a switch will always occur after three same-language trials. Therefore, dummy items (unused pictures from the picture naming experiments) were added to create longer same language runs within the lists. Three lists were made with the paired blocks and rotated using a Latin square creating 12 experimental lists. Each participant saw one list.

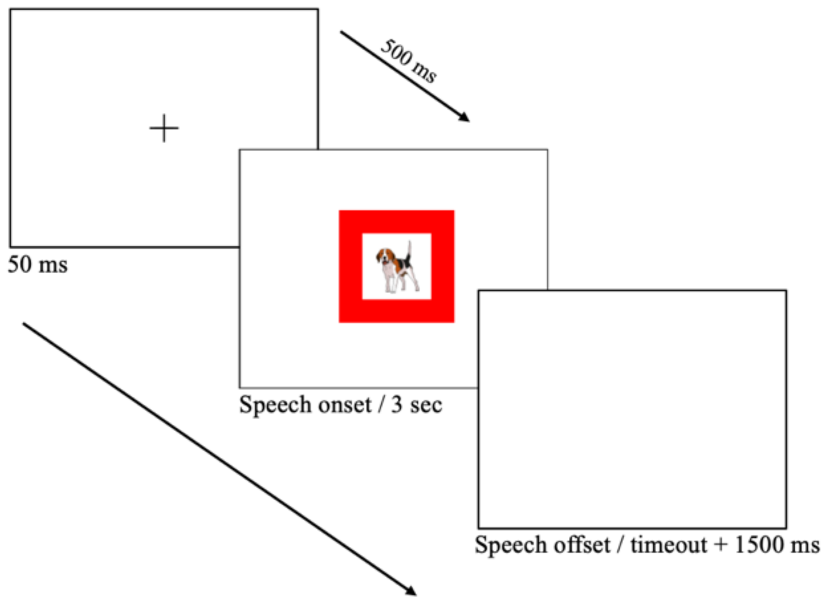
Naming language was indicated by the colour of a 500 x 500 pixel frame around the picture to be named. For half of the participants blue and green frames indicated L1 naming, and red and yellow indicated L2 naming. The colour-language assignment was counterbalanced for the other half. Two colours were used for each language to avoid confounding cue switching with language switching (e.g., Heikoop et al., 2016). This way colour would switch on all trials, while language would switch only on some trials.

Procedure

Participants completed the language switching experiment at the start of the third and final session. Colour-language assignment was shown on a handout, first just the frames with naming language written underneath each frame and then frames with example dummy pictures and the correct word written below. Before the experiment proper there were two practice blocks with 15 trials each. The 15 most frequent cognates and noncognates in the dummy items were used in the practice block. The practice blocks had a similar switching structure to the experimental blocks. In each practice block, approximately half of the items were cognates and half were noncognates.

There were eight experimental blocks. As dummy strings were included in each block, block length varied from 30 to 39 trials (M = 33.6). There was a pause between each block. The trial structure is visualised in Figure 12. Each trial proceeded in the same way as in the picture naming task, beginning with the presentation of a 50 ms beep and fixation cross placed at the centre of the screen. After 500 ms the picture, with a coloured frame cueing naming language, was presented. The language cue and the picture to be named appeared at the same time, to maximise measurements of the switching effects (C. Li & Gollan, 2018). Speech onset and speech offset were detected by the voice key. The picture disappeared from the screen at speech onset or at timeout 3000 ms after picture onset if the participant failed to respond. The next trial began 1500 ms after speech offset (or timeout). Participants were instructed to respond as quickly and accurately as possible.

Figure 12
Switching task trial structure



5.3.3 Flanker task

Design and materials

In this task, the participants indicated whether the central arrow in a visually presented string of 5 arrows pointed left or right. Stimuli and design were the same as in Zhou and Krott (2018). The items were four .png files with a 1024 x 768 resolution. The arrows were black and appeared on a white background. In congruent images all arrows pointed in the same direction (left or right) and in incongruent images the surrounding arrows all pointed in the opposite direction of the central arrow.

Procedure

The flanker task was completed on the third day after the language switching task. There was a practice block with 24 trials and two experimental lists of 96 trials. Stimuli were presented in the same order for all participants. Participants were asked to focus on the arrow in the middle and ignore the others. They were instructed to press the z key on the keyboard if the middle arrow pointed left and the m key if it pointed right. They were instructed to respond as quickly and accurately as possible. The participants were seated in front of the screen with their left index finger placed on the z key and the right on the m key. Each trial began with a fixation cross presented for 800 ms. After a 500 ms window with a blank screen the stimulus was presented. The string of five arrows could appear either above or below the fixation cross. After the participants responded the next trial began after a 500 ms blank screen. The response timeout was at 5000 ms.

5.4 Results

In this section the results from picture naming and the language switching task are presented separately. The data from the picture naming sessions in Norwegian and English are combined for the analysis. The results from the flanker task, as well as the *General L2 proficiency* and *L2 accent and interest* factors were added as predictors to models of reaction times and accuracy in the picture naming and language switching experiments. The fixed effects were cognate status, naming language, and picture naming session. The latter refers to whether the picture naming data were produced during the first or the second picture naming session, irrespective of whether the order was L1-L2 or L2-L1. In the language switching experiment the fixed effects additionally included trial type (i.e., switch or stay). As described in the methods section, the participants completed two picture naming sessions, one in English and one in Norwegian. The order was counterbalanced in order to avoid a systematic bias towards one language, but one cannot rule out that naming pictures in L1 in the first session or L2 in the first session could have different effects on naming behaviour in the following experimental sessions, especially for unbalanced bilinguals. Block order effects have previously been observed when testing different languages on the same day (Gollan et al., 2008; Misra et al., 2012), therefore session was included in the models to investigate if the order of naming language (whether the first session was in English or Norwegian) could have influenced the reaction times. All data processing and statistical analysis of the data was conducted in R (R Core Team, 2022). Models of reaction times and accuracy were fitted using the package lme4 (version 1.1.28; Bates et al., 2015). The alpha for the study was set at 0.05, apart from model reduction where the alpha was set at 0.1 for retaining model terms. Figures were made with the packages ggplot2 (version 3.4.2; Wickham, 2016) and cowplot (version 1.1.1; Wilke, 2020).

5.4.1 Picture naming

Reaction times (RTs) and error rates were collected from the picture naming tasks. Errors in naming (including disfluencies, hesitations, and failures to respond) were registered by the experimenter (1429 errors in total, 43691 correct)[8]. The data were screened for short latencies (≤ 400 ms, $N = 55$), all of which stemmed from error trials or recording failures, and these were included in the error count. Error removal resulted in a 3.17% loss of data (L1: 2.90%, L2: 3.44%). Before modelling the data, trials following errors were also removed, bringing the data loss to 6.01%. Visual inspection of the correct data showed that the RTs were positively skewed, which is not uncommon for these kinds of data (Baayen & Milin, 2010). Following Baayen and Milin, the data were modelled with minimal a-priori data trimming. Observations exceeding 4 standard deviations from the by-participant and by-language mean were removed (English: 0.79%, Norwegian: 0.83%) bringing the total data loss to 6.77%.

Analysis of reaction times

A linear mixed effects model was fitted with naming language (English vs. Norwegian), cognate status (Cognate vs. Noncognate), and picture naming session (Session 1 or Session 2) as fixed effects, coded at -0.5 and 0.5 (-0.5 for the first level listed in the parentheses). The model included random intercepts for subject (cognate status and naming language random slopes) and item (naming language and session random slopes), corresponding to the maximal random effects structure justified by the design (see Barr et al., 2013). All interactions were included for both the fixed and random effects. The *General L2 proficiency* and *L2 accent and interest* factors found in Chapter 4 and the difference between congruent and incongruent trials on the flanker task (described in Section 5.3.3) were centred and added as predictors, including two- and three-way interactions with naming language and cognate status. The models were run with the bobyqa optimiser.

The model was run with different RT transformations (log, log10, inverse). The assumptions of linearity and normality of the residuals were to some extent

[8] In the stimulus set two pictures were included as noncognates (L2 *maze* – L1 *labyrint* and L2 *bull* – L1 *okse*) based on their modal names given in the MultiPic database. Both words have English synonyms that are in fact cognates between Norwegian and English (*labyrinth* and *ox*, respectively) and therefore these were excluded from the data. The final model was run both with and without these items, it did not change the results. 960 observations were removed in total.

violated in all, but the best results were obtained with the inverse transformation. Marginal R^2 (0.059) and conditional R^2 (0.479) were also the highest with the inverse transformation. With the inverse transformation the residual plot indicated no issues with heteroskedasticity and VIF scores indicated no issues with multicollinearity.

As the residuals were in violation with assumptions of linearity and normality, observations that were the source for residuals more than 2.5 standard deviations from the mean were removed (Baayen & Milin, 2010, p. 10). This resulted in an additional 2.06% data loss, bringing the total data loss to 8.69%. This improved the model fit, and the linearity and normality of the residuals, albeit not for the extreme ends of the distribution. A final check of the observations that were the source of the residuals at the extreme ends did not reveal any specific patterns[9] and the data were not trimmed any further[10]. The final dataset is visualised in Figure 13.

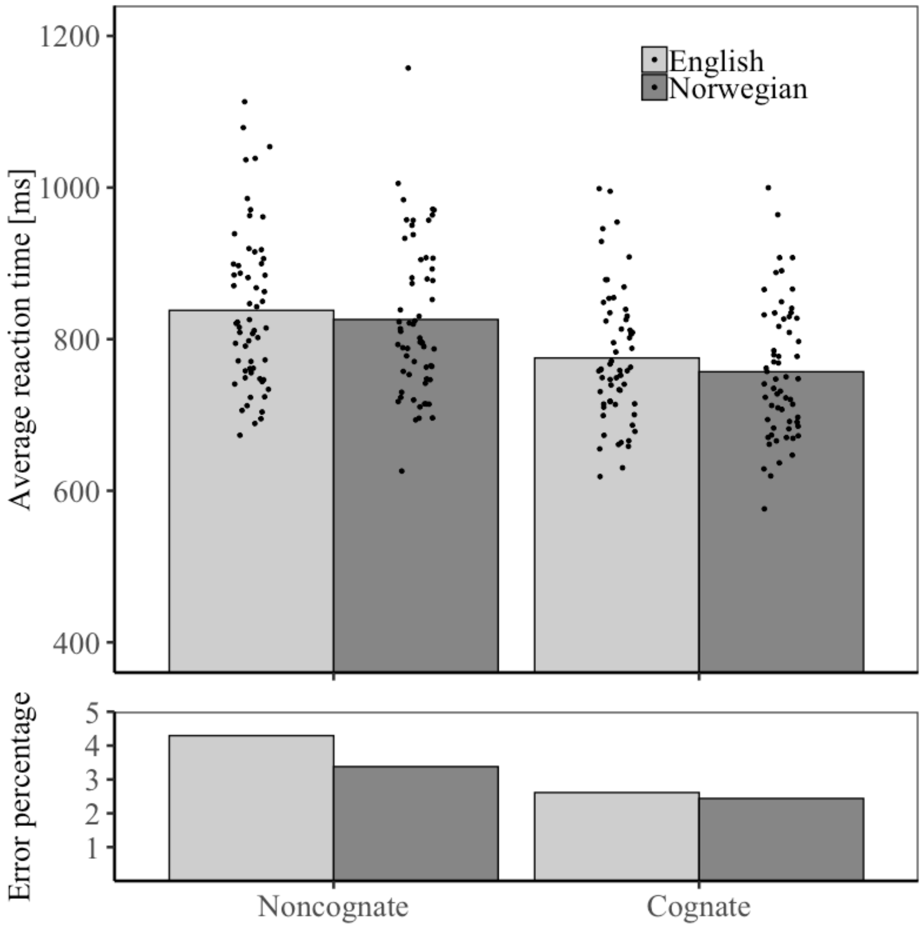
The fixed effects structure of the model was evaluated using the drop1 function[11]. The *General L2 proficiency* predictor and all its interactions were removed. For the flanker task predictor, the three-way interaction with cognate status and naming language, as well as the two-way interaction with naming language was removed. The final model fit had the highest marginal R^2 (0.064) and conditional R^2 (0.540). The final model and the results are reported in Table 12 and residual plots are printed in Appendix J, Figure J1. Model comparisons with likelihood ratio tests were used to test the significance of fixed effects.

[9] All participants contributed at least one observation, the data points were relatively equally distributed across items (86 English words and 89 Norwegian words), as well as across language and cognate status (145 Norwegian cognates, 130 Norwegian noncognates, 128 English cognates and 126 English noncognates).

[10] I.e. minimal a-priori trimming and visual inspection of model residuals to further improve the model (Baayen & Milin, 2010). Since several studies cited here have used a +/- 3 SD trim, and the model fit with the +/- 4 SD trim was less reliable for the residuals at the extreme ends, I also tried removing +/- 3 SD. This did not improve the model fit, nor change the results.

[11] A post-hoc check also showed that the final model corresponded to the smallest Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC) – see Appendix K.

Figure 13
Mean reaction times and error rates by naming language and cognate status in the picture naming task



Note. Dots correspond to individual participant means.

The fixed effects cognate status and naming language were both significant. Pictures with cognate names were named faster (M = 766.10 ms, SD = 165.18 ms) than noncognate pictures (M = 831.92 ms, SD = 202.74 ms) and naming was faster in L1 Norwegian (M = 790.34 ms, SD = 182.18 ms) compared to L2 English (M = 805.29 ms, SD = 191.71 ms). There was no significant interaction between cognate status and language. The mean difference in RTs between cognate and noncognate cognate words was 62.85 ms in English and 68.89 ms in Norwegian.

Table 12
Model output for picture naming reaction times

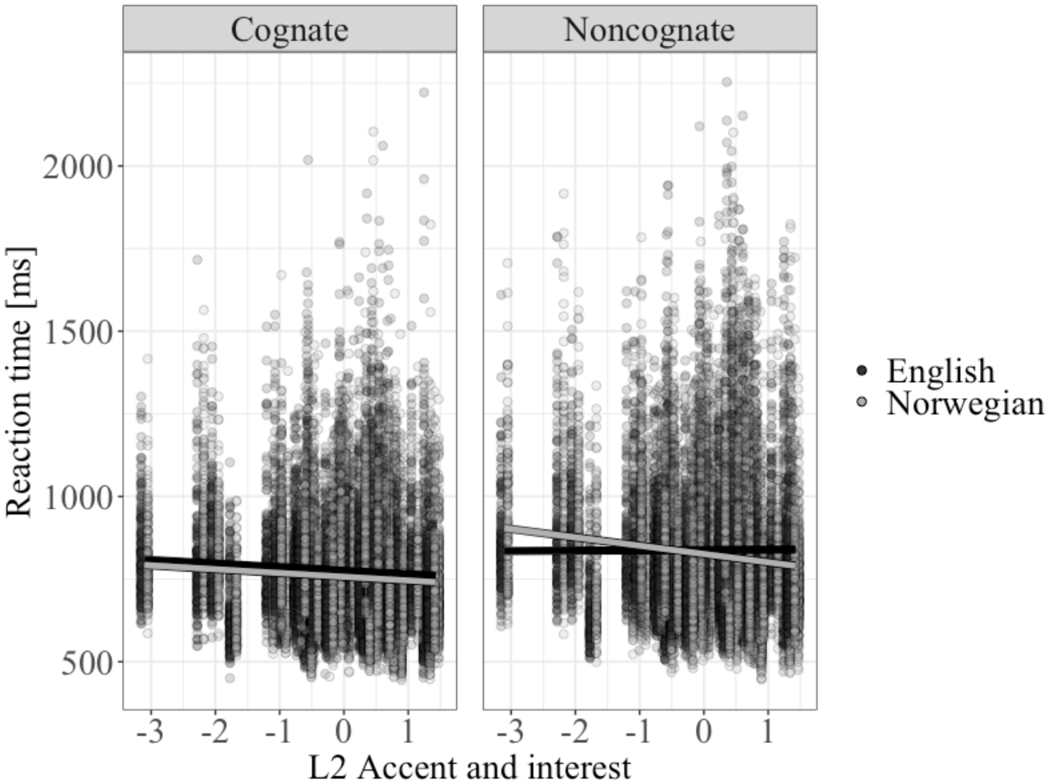
Model: -1000 *1/RT ~ CognateStatus*NamingLanguage*Session + cAccent *CognateStatus*NamingLanguage + cFlanker_diff*CognateStatus+ (CognateStatus + NamingLanguage + CognateStatus:NamingLanguage|SUBJ_ID) + (NamingLanguage+Session + NamingLanguage:Session |Item)

	Model summary			Model comparison		
	Estimate	*SE*	*t*	χ^2	df	*p*
Fixed effects						
(Intercept)	-1.3046	0.0206	-63.376			
CogStatus	**0.0977**	**0.0242**	**4.041**	**25.88**	**1**	**<.001**
Language	**-0.0301**	**0.0125**	**-2.417**	**5.73**	**1**	**.017**
Session	-0.0061	0.0108	-0.572	0.003	1	.957
cAccent	**-0.0123**	**0.0195**	**-0.632**	**4.17**	**1**	**.041**
cFlanker_diff	0.0001	0.0002	0.489	1.25	1	.264
CogStatus:Language	0.0116	0.0167	0.694	0.16	1	.692
CogStatus:Session	**0.0432**	**0.0101**	**4.270**	**16.21**	**1**	**<.001**
Language:Session	**0.1444**	**0.0697**	**2.073**	**4.81**	**1**	**.028**
CogStatus:cAccent	0.0023	0.0039	0.586	0.76	1	.384
Language:cAccent	**-0.0209**	**0.0111**	**-1.871**	**4.71**	**1**	**.030**
CogStatus:cFlanker_diff	**0.0001**	**0.00004**	**2.830**	**7.56**	**1**	**.006**
CogStatus:Language:Session	**0.0310**	**0.0171**	**1.810**	**9.43**	**1**	**.009**
CogStatus:Language:cAccent	**-0.0183**	**0.0066**	**-2.781**	**7.26**	**1**	**.007**
Random effects	s^2					
Item (Intercept)	0.014					
Subject (Intercept)	0.018					

Note. CogStatus = Cognate status (Cognate vs. Noncognate); Language = Naming language (English vs. Norwegian); Session (Session 1 vs. Session 2); cAccent (centred factor scores from the *L2 accent and interest* factor); cFlanker_diff (centred RT differences between congruent and incongruent trials on the flanker task). Significant results in bold.

The *L2 accent and interest* predictor main effect was significant and involved in a significant two-way interaction with naming language, as well as a significant three-way interaction with cognate status and naming language. Figure 14 shows that naming of cognates, in both languages, and L1 noncognates speeds up as the value of the predictor increases, while L2 noncognates appear to be largely unaffected.

Figure 14
Three-way interaction between naming language, cognate status, and the L2 accent and interest predictor in the picture naming experiment



Note. *L2 accent and interest* factor scores are centred.

Submodels[12] separated by language showed that in English there is a significant interaction between cognate status and the predictor (Estimate = 0.017, *SE* = 0.005, $t = 3.59$, $p < .001$). While cognates were named more quickly as the predictor value increased (trend: -0.008, *SE* = 0.0196), noncognates showed an effect in the opposite direction (trend: 0.004, *SE* = 0.0199). In Norwegian there is a marginal interaction between cognate status and the predictor (Estimate = -0.014, *SE* = 0.007, $t = -1.96$, $p = .051$). Both cognates and noncognates are named more quickly when the predictor value increases, but the effect is stronger for noncognates (trend: -0.026, *SE* = 0.0197) compared to cognates (trend: -0.019, *SE* = 0.0225).

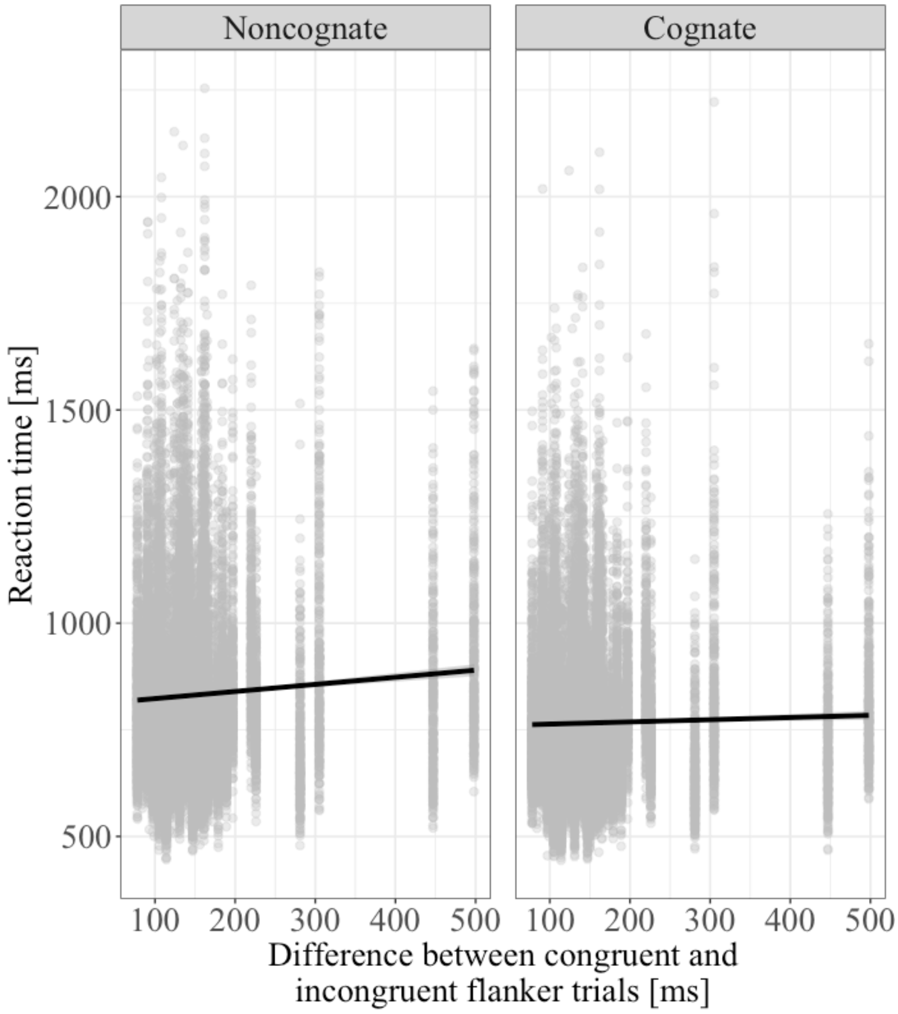
There was no main effect of flanker task performance on RTs in the picture naming experiment, but it appeared in a significant interaction with cognate status. This is

[12] -1000 * 1/RT ~ CognateStatus*cAccent +
(CognateStatus|SUBJ_ID) + (1|Item)

visualised in Figure 15. Larger RT differences between congruent and incongruent trials in the flanker task are associated with longer response times in the picture naming task for both cognates and noncognates, but the effect is larger for noncognates (trend: 0.00017, *SE* = 0.00021) compared to cognates (trend: 0.00004, *SE* = 0.00022).

Figure 15

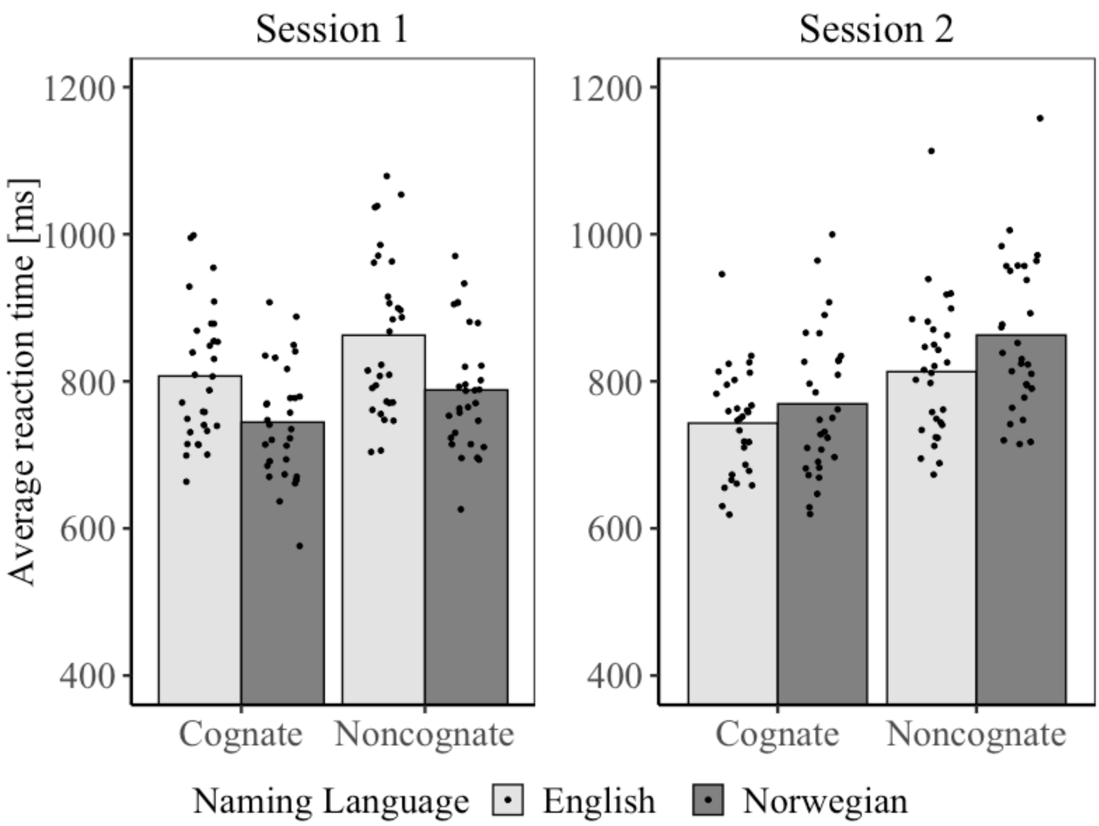
Plot of interaction between cognate status and the flanker task predictor in the picture naming task



There was no significant main effect of picture naming session, but it was involved in significant two-way interactions with both cognate status and naming language, as well as a three-way interaction with cognate and naming language. This three-way interaction is visualised in Figure 16 and was further analysed using nested submodels, one for each of the two-way interactions (CognateStatus:Session and NamingLanguage:Session)[13].

Figure 16

Plot of reaction times in the first and second picture naming session, grouped by cognate status and naming language



Note. Separate plots for RTs obtained in the first picture naming session (Session 1) and the second picture naming session (Session 2). Dots correspond to individual participant means.

[13] Cognate status and Session model:
-1000 * 1/RT ~ CognateStatus/Session + (CognateStatus|SUBJ_ID) + (NamingLanguage|Item)
Naming language and session model:
-1000 * 1/RT ~ NamingLanguage/Session + (CognateStatus|SUBJ_ID) + (NamingLanguage|Item)

The analysis revealed that cognates were named significantly faster than noncognates in both sessions, however the reaction times are differentially affected in the first session compared to the second session. Cognates were on average named 19.63 ms faster in the second session compared to the first (Estimate = -0.025, *SE* = 0.0026, t = -9.77, p <.001), while noncognates were on average named 12.85 ms more slowly (Estimate = 0.023, *SE* = 0.0027, t = 8.84, p <.001). The cognate words named in the second session have been more frequently and recently activated (when being named in the other language) than noncognate words, and the shorter naming latencies for cognates could reflect increased activation cross-linguistically. Turning to the overall effect of naming language, L1 naming was faster than L2 naming, in line with previous studies. However, L1 naming was slower in the second session compared to the first session (Estimate = 0.095, *SE* = 0.0326, t = 2.93, p = .005), while L2 naming was faster when completed in the second session compared to the first session (Estimate = -0.099, *SE* = 0.0326, t = -3.03, p = .004). L1 naming was on average 68.30 ms faster than L2 naming in the first session, but this reversed in the second session where L2 naming was 38.06 ms faster than L1 naming on average.

Analysis of error rates

Error rates were analysed with generalised mixed effects modelling and the bobyqa optimiser[14]. There were significantly fewer errors in Norwegian than English (*M* L1 = 2.9%; *M* L2 = 3.4%; Estimate = -0.225, *SE* = 0.082, $\chi^2(1) = 6.08$, p = .014) and fewer errors in the second picture naming session compared to the first (*M* Session 1 = 3.81%; *M* Session 2 = 2.52%; Estimate = -0.511, *SE* = 0.080, $\chi^2(1)$ = 22.11, p < .001). There was a significant main effect of the flanker predictor. Fewer errors were linked to a smaller difference between congruent and incongruent flanker trials (Estimate = 2.795, *SE* = 1.190, $\chi^2(1)$ = 4.13, p = .042). There were also fewer errors on cognate trials than noncognate trials (Cognate *M* = 2.53%; Noncognate *M* = 3.84%; Estimate = -0.650, *SE* = 0.213, $\chi^2(1)$ = 8.21, p = .004).

[14] The full model did not converge and there were issues with the scale of the flanker predictor. The flanker predictor was rescaled, the random structure was reduced and then evaluated through model comparisons using ANOVAs. A post-hoc check also showed that the final model corresponded to the smallest Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC) – see Appendix K.

Final model: error ~ CognateStatus + Session + cAccent + NamingLanguage + I(cFlanker_diff/1000) +
CognateStatus:Session + cAccent:NamingLanguage +
I(cFlanker_diff/1000):NamingLanguage + (NamingLanguage|SUBJ_ID) + (1|Item)

Finally, the *L2 accent and interest* predictor main effect was not significant (Estimate = 0.059, SE = 0.101, $\chi^2(1)$ = 0.66, p = .415). The full result table is printed in Appendix L, Table L1.

There was a significant interaction between cognate status and picture naming session (Estimate = -0.265, SE = 0.114, $\chi^2(1)$ = 5.24, p = .022). Both types of words were named with fewer errors in the second session, but the difference between noncognate and cognate trials (i.e., the cognate benefit) was larger in the second session (mean difference in Session 1 = 1.21% vs. Session 2 = 1.39%). The results also show a significant interaction between naming language and the *L2 accent and interest* predictor (Estimate = -0.221, SE = .090, $\chi^2(1)$ = 5.61, p = .018). However, the effect of the predictor was not significant in the either of the language-specific submodels[15].

5.4.2 Language switching

As for the picture naming experiment, reaction times and error rates were collected from the language switching experiment. Errors (including disfluencies, hesitations, and failures to respond) were registered by the experimenter during the experiment. In addition, registered response times under 500 ms (N = 3) were excluded, removing 560 trials in total and leaving 10720 correct trials[16]. All response times below 500 ms were disfluencies that were not registered during the experiment. Error removal resulted in a 4.96% data loss (L1: 4.72%; L2: 5.21%). Errors and trials following errors were removed before analysing the reaction times (data loss: 9.50%) and all observations exceeding 4 standard deviations from the by-participant and by-language mean (L1: 0.20%, L2: 0.34%) were excluded as well, bringing the total data loss to 9.74%.

Analysis of reaction times

As for the picture naming task, the data were analysed using linear mixed effects modelling with the bobyqa optimiser. The data were first fit to a maximal model with the same terms as the picture naming analysis, as well as the same data trim

[15] Model:
error ~ I(cFlanker_diff/1000) + (1|SUBJ_ID) + (1|Item)
English result: Estimate = 0.127, SE = 0.115, z = 1.10, p =.271.
Norwegian result: Estimate = 0.065, SE = 0.122, z = 0.53, p = .596.

[16] Prior to the analysis all 240 instances of item 570 were removed (because *bull* was erroneously categorised as a noncognate word, described in more detail for the picture naming dataset).

and transformation (+/- 4 *SD* and inverse, respectively). In addition, trial type (Stay vs. Switch) was added as a fixed effect coded at -0.5 and 0.5, in interaction with the predictors, and in random slopes for subject and item. The full model did not converge. The random structure was reduced incrementally, leaving a model with random intercepts for item and subject, as well as random slopes for cognate status, naming language and trial type by subject. The flanker score was rescaled for both the latency and error analysis.

Visual inspection of the residuals revealed no issues with heteroscedasticity and VIF scores indicated no issues with multicollinearity. The residuals were in violation with assumptions of normality, and to a lesser degree linearity, and observations that were the source for residuals more than 2.5 standard deviations from the mean were removed (Baayen & Milin, 2010, p. 10). This resulted in an additional 1.43% data loss (146 data points removed), bringing the total data loss to 11.04%. This improved the model fit and the distribution of the residuals. As in the picture naming task, there were tails at the extreme ends of the distribution. Further inspection of the observations that were the source of the residuals did not reveal any specific patterns[17] and the data were not trimmed any further. The final dataset is visualised in Figure 17.

The model terms were evaluated using the drop1 function[18]. All four-way interactions were removed, as well as all interactions involving the flanker predictor. In addition, 6 three-way interactions and 3 two-way interactions were removed. Table 13 shows the final model and the results. Model comparisons with likelihood ratio tests were used to test the significance of fixed effects. The final residual plot is printed in Appendix J, Figure J2.

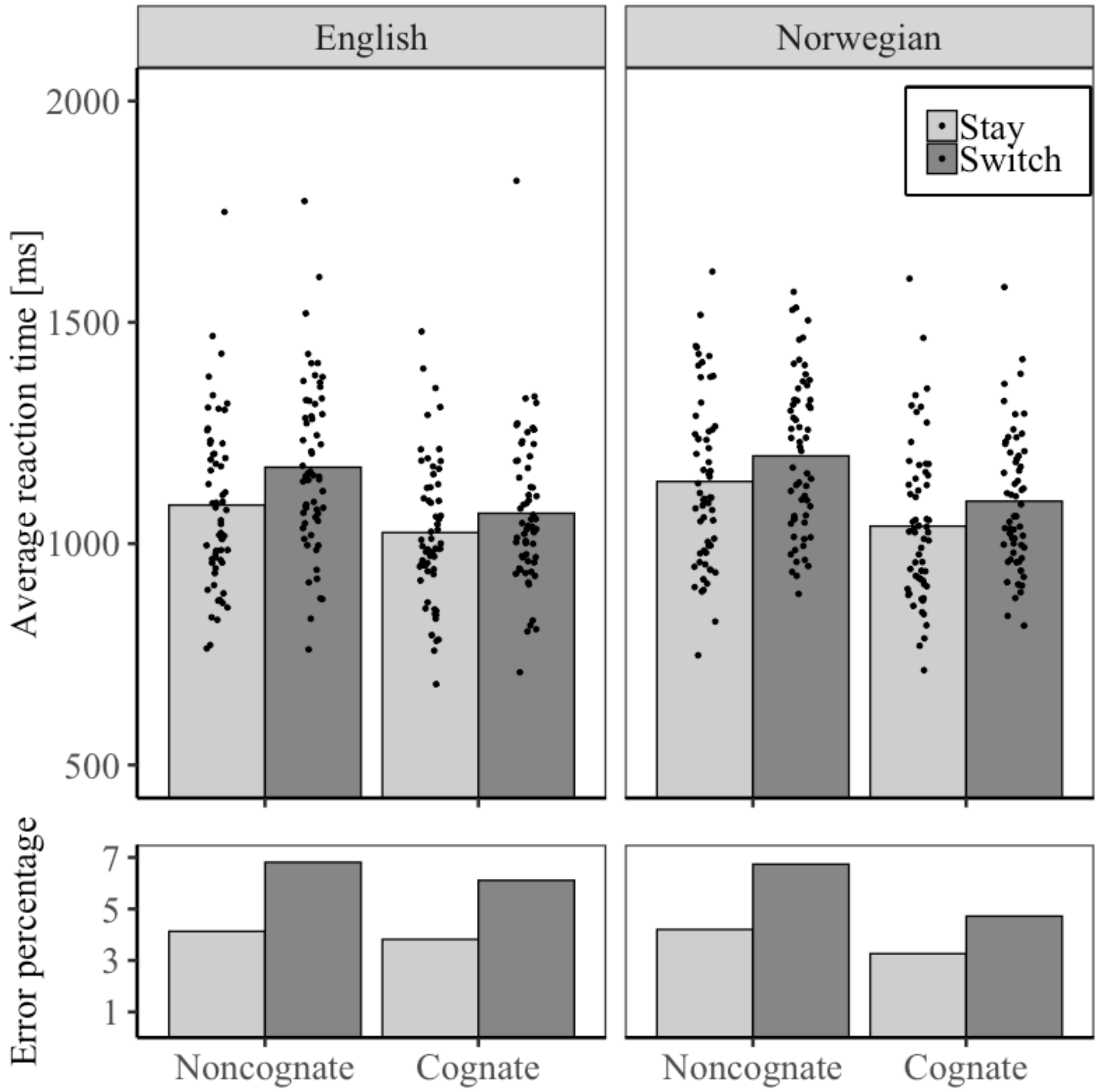
[17] Data points in the tails stemmed from more than half (N = 38) of the participants and words named in both English (N = 20) and Norwegian (N = 27). Number of observations in the tails by cognate status, naming language and trial type provided in the overview below.

Cognates			Noncognates		
Language	Trial type	N	Language	Trial type	N
English	Stay	10	English	Stay	9
English	Switch	3	English	Switch	5
Norwegian	Stay	14	Norwegian	Stay	14
Norwegian	Switch	5	Norwegian	Switch	6

[18] A post-hoc check also showed that the final model corresponded to the smallest Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC) – see Appendix K

Figure 17

Mean reaction times and error rates by naming language, trial type and cognate status in the language switching task



Note. Dots correspond to individual participant means.

Table 13
Model output from the analysis of reaction times in the language switching task

Model: -1000 *1/RT ~ CognateStatus + TrialType + NamingLanguage + cL2prof + cAccent + Session + I(cFlanker_diff/1000) + CognateStatus:TrialType + CognateStatus:NamingLanguage + TrialType:NamingLanguage + CognateStatus:TrialType:NamingLanguage + cL2prof:CognateStatus + cL2prof:NamingLanguage + cL2prof:CognateStatus:NamingLanguage + cAccent:NamingLanguage + cAccent:TrialType + cAccent:NamingLanguage:TrialType + Session:CognateStatus + Session:TrialType + TrialType:CognateStatus:Session + (CognateStatus + NamingLanguage + TrialType |SUBJ_ID) + (1|Item)

	Model summary			Model comparison		
	Estimate	*SE*	*t*	χ^2	df	*p*
Fixed effects						
(Intercept)	-0.9632	0.0196	-49.056			
CogStatus	**0.0700**	**0.0215**	**3.260**	**9.63**	**1**	**.002**
TrialType	**0.0643**	**0.0042**	**15.458**	**93.77**	**1**	**<.001**
Language	**0.0242**	**0.0072**	**3.360**	**9.76**	**1**	**.002**
cL2prof	-0.0099	0.0158	-0.631	0.54	1	.462
cAccent	-0.0309	0.0190	-1.627	1.58	1	.209
Session1	-0.0134	0.0073	-1.837	2.11	1	.146
cFlanker_diff	**0.5493**	**0.2032**	**2.703**	**6.83**	**1**	**.009**
CogStatus:TrialType	-0.0130	0.0070	1.856	3.47	1	.063
CogStatus:Language	**0.0199**	**0.0070**	**2.843**	**8.35**	**1**	**.004**
TrialType:Language	-0.0003	0.0070	-0.037	0.0001	1	.992
CogStatus:cL2prof	0.0047	0.0053	0.884	0.78	1	.377
Language:cL2prof	-0.0043	0.0076	-0.564	0.34	1	.561
Language:cAccent	-0.0133	0.0084	-1.574	2.48	1	.115
TrialType:cAccent	0.0079	0.0046	1.697	2.84	1	.092
CogStatus:Session	-0.0087	0.0070	-1.237	1.29	1	.256
TrialType:Session	0.0141	0.0073	1.950	1.99	1	.158
CogStatus:TrialType:Language	-0.0243	0.0140	-1.742	3.04	1	.081
CogStatus:Language:cL2prof	0.0134	0.0072	1.859	3.46	1	.063
TrialType:Language:cAccent	0.0156	0.0081	1.919	3.68	1	.055
CogStatus:TrialType:Session	**0.0311**	**0.0140**	**2.231**	**4.98**	**1**	**.026**
Random effects	s^2					
Item (Intercept)	0.005					
Subject (Intercept)	0.017					

Note. CogStatus = Cognate status (Cognate vs. Noncognate); Trial type (Stay vs. Switch); Language = Naming language (English vs. Norwegian); Session (Session 1 vs. Session 2); cAccent (centred factor scores from the *L2 accent and interest* factor); cL2prof (centred factor scores from the *General L2 proficiency* factor); cFlanker_diff (centred and rescaled RT differences between congruent and incongruent trials on the flanker task). Significant results in bold.

The fixed effects cognate status, naming language, trial type (switch or stay) and performance on the flanker task were all significant. Cognates were named faster than noncognates (Cognate: M = 1057.0 ms, SD = 274.86 ms, Noncognate: M = 1148.82 ms, SD = 319.56 ms), naming was faster in English than in Norwegian (English: M = 1086.54 ms, SD = 299.37 ms, Norwegian: M = 1116.17 ms, SD = 301.56 ms) and participants were faster on stay trials compared to switch trials (Stay: M = 1071.71 ms, SD = 302.03 ms, Switch: M = 1132.08 ms, SD = 296.49 ms). Finally, a smaller difference between congruent and incongruent trials on the flanker task was linked to faster naming. There was a significant two-way interaction between cognate status and naming language. L2 cognates were on average named 82.43 ms faster than noncognates (Cognates: M = 1046.49, SD = 271.83, Noncognates: M = 1128.92, SD = 320.64), while the average cognate facilitation effect was 101.41 ms in L1 (Cognates: M = 1067.34, SD = 277.47, Noncognates: M = 1168.75, SD = 317.29). A nested model[19] found that the difference in reaction times between the two languages was significant for noncognate words (M = 39.83, Estimate = 0.034, SE = 0.0099, t = 3.467, p = .001), but not for cognate words (M = 20.86, Estimate = 0.014, SE = 0.0096, t = 1.496, p = .142). The cognate benefit appears to be greater for L1 Norwegian, although the submodel[20] shows an effect for both languages. (L1: Estimate = -0.080, SE = 0.023, t = -3.508, p < .001; L2: Estimate = -0.061, SE = 0.022, t = -2.734, p = .009). The three-way interaction between cognate status, naming language, and trial type did not reach significance.

The results show a significant three-way interaction between cognate status, trial type and whether a language was named in the first or second picture naming session. This interaction is visualised in Figure 18. Mean switch costs were similar for cognates (M = 55.52 ms) and noncognates (M = 59.84 ms) for responses in the language participants used in the first session of the simple picture naming task. For responses produced in the language used in the second picture naming session, however, switch costs for cognates are slightly reduced (M = 44.59 ms) while there is a larger switch cost for noncognates (M = 83.78 ms). Submodels divided by trial

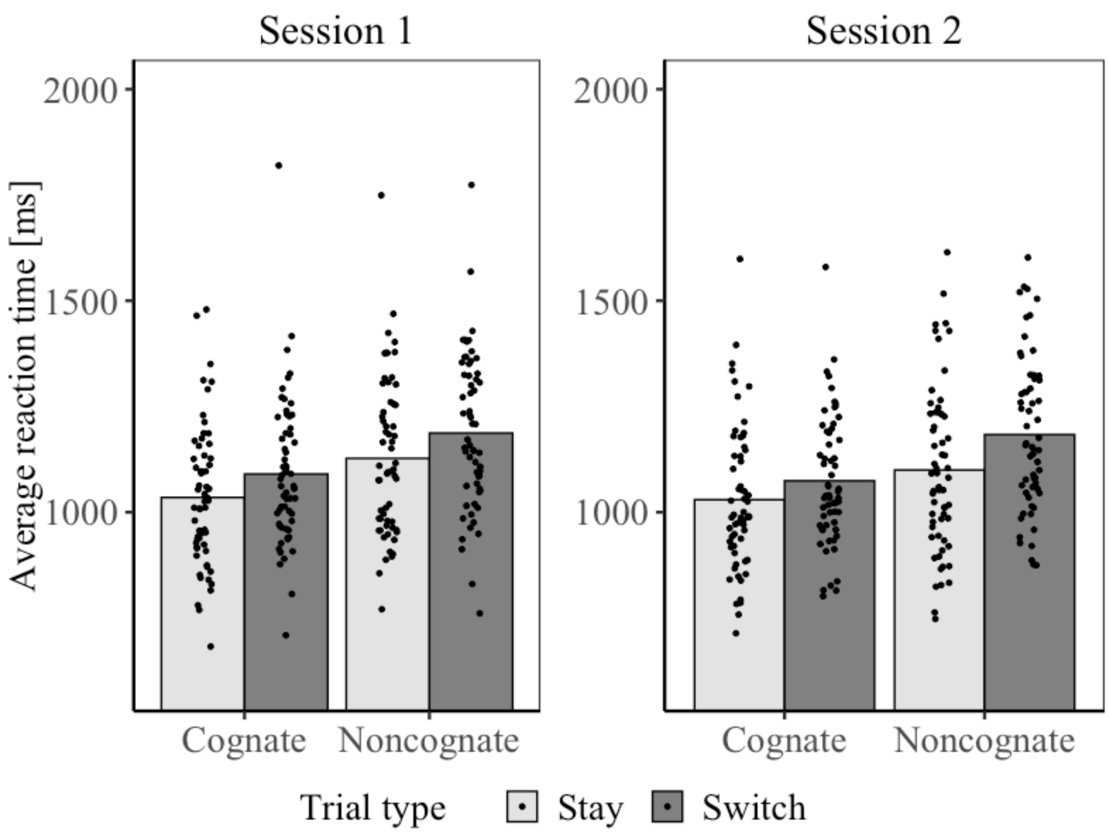
[19] -1000 * 1/RT ~ CognateStatus/NamingLanguage + (CognateStatus|SUBJ_ID) + (NamingLanguage|Item)

[20] -1000 * 1/RT ~ NamingLanguage/CognateStatus + (CognateStatus|SUBJ_ID) + (NamingLanguage|Item)

type[21] show that there is no significant interaction between noncognates and session (Estimate = -0.003, *SE* = 0.0069, t = -0.442, p = .659), nor cognates and session (Estimate = -0.010, *SE* = 0.0067, t = -1.490, p = .136) on switch trials. However, on stay trials there is a significant interaction between noncognates and session (Estimate = -0.029, *SE* = 0.0074, t = -3.832, p <.001). The interaction between cognate status and session remains insignificant (Estimate = -0.005, *SE* = 0.0072, t = -0.726, p = .468).

Figure 18

Plot of mean reaction times in the switching task grouped by cognate status, trial type and picture naming session



Note. Separate plots for RTs produced in the language used in the first simple picture naming session (Session 1) and the language used in the second simple picture naming session (Session 2). Dots correspond to individual participant means.

[21]The following model was run separately on RTs from stay trials and switch trials:
-1000 * 1/RT ~ CognateStatus/Session + (CognateStatus|SUBJ_ID) + (1|Item)

Finally, there were two marginal 3-way interactions (p <.07) involving the bilingual profile predictors. The first was a marginal interaction between cognate status, naming language, and the *General L2 proficiency* predictor. In general, naming latencies are shorter with higher values of the predictor, supported by the negative trend found for cognates and noncognates in both Norwegian and English. The effect of the predictor appears to be larger for L1 cognates (trend: -0.0178, SE = 0.0158) compared to the other conditions (L1 noncognates trend: -0.0064, SE = 0.0157; L2 cognates trend: -0.0068, SE = 0.0174; L2 noncognates trend: -0.0088, SE = 0.0171). The second marginal interaction was between trial type, naming language, and the *L2 accent and interest* predictor. As for the other predictor, the general pattern showed shorter naming latencies with higher values of the predictor. The effect appears to be stronger in Norwegian (Stay trend: -0.0453, SE = 0.0199; Switch trend: -0.0297, SE = 0.0179), especially on stay trials, compared to trials with naming in English (Stay trend: -0.0243, SE = 0.0213; Switch trend: -0.0242, SE = 0.0194).

Analysis of error rates

Error rates were analysed with generalised mixed effects modelling using the bobyqa optimiser[22]. There were more errors on switch trials than on stay trials (M = 6.08% vs. 3.85%, Estimate = 0.531, SE = 0.091, $\chi^2(1) = 32.62$, $p < .001$) and more errors when naming pictures in the language spoken in the first experimental session compared to the second (M = 5.46% vs. 4.47%, Estimate = -0.209, SE = 0.089, $\chi^2(1) = 5.98$, $p = .014$). Finally, a larger RT difference between congruent and incongruent trials on the flanker task were associated with higher error rates on the switching task (Estimate = 2.660, SE = 1.324, $\chi^2(1) = 3.84$, $p = .050$).

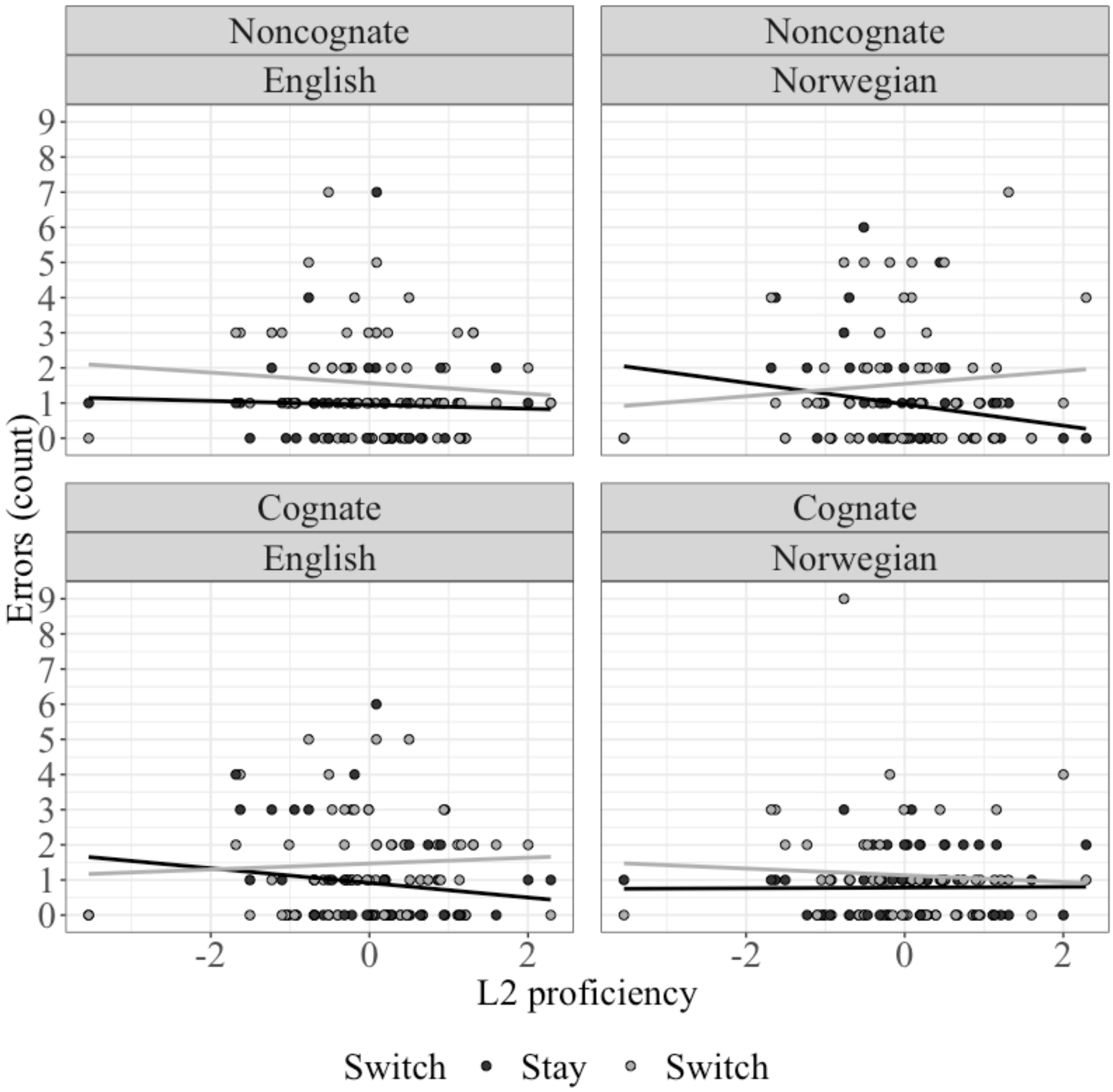
[22] The full model did not converge. The random structure was reduced to a random intercept only for item and a random intercept for subject with a random slope for cognate status. Model reduction and model evaluation were both done using ANOVAs. A post-hoc check also showed that the final model corresponded to the smallest Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC) – see Appendix K.

Final model:
error ~ CognateStatus + NamingLanguage + TrialType + Session + cL2prof + I(cFlanker_diff/1000) + CognateStatus:Session + cL2prof:CognateStatus + cL2prof:NamingLanguage + cL2prof:TrialType + CognateStatus:NamingLanguage + CognateStatus:TrialType + TrialType:NamingLanguage + CognateStatus:NamingLanguage:TrialType + cL2prof:NamingLanguage:TrialType + CognateStatus:cL2prof:TrialType + CognateStatus:NamingLanguage:cL2prof + cL2prof:CognateStatus:NamingLanguage:TrialType + (CognateStatus|SUBJ_ID) + (1|Item)

The only significant interaction was a four-way interaction between cognate status, naming language, trial type and the *General L2 proficiency* predictor (Estimate = -1.072, $SE = 0.400$, $\chi^2(1) = 7.00$, $p = .008$), visualised in Figure 19. There were two marginally significant interactions between cognate status and session (Estimate = 0.354, $SE = 0.179$, $\chi^2(1) = 3.67$, $p = .055$) and L2 proficiency and trial type (Estimate = 0.173, $SE = 0.100$, $\chi^2(1) = 3.49$, $p = .062$). The full table is printed in Appendix L, Table L2.

Figure 19

Plot of the four-way interaction between cognate status, naming language, trial type and General L2 proficiency in the accuracy analysis of the language switching task



Note. The figure shows the number of errors for different factor scores of the General L2 proficiency factor (L2 proficiency). The results are grouped by cognate status and naming language. Trial type is indicated by colour: black for stay trials and grey for switch trials.

Figure 19 suggests that higher *General L2 proficiency* factor scores are generally associated with lower error rates, except for on switch trials for L2 cognates and L1 noncognates. Language specific submodels showed that trial type was significant for both languages. There were more errors on switch trials than on stay trials in both English (M = 6.45% vs. 3.97%) and Norwegian (M = 5.71% vs. 3.72%). This was the only significant effect in the model of the English error data. In Norwegian there was additionally a significant 3-way interaction between cognate status, trial type and the *General L2 proficiency* predictor (Estimate = -0.643, SE = 0.279, z = -2.310, p = .021). Further analyses did not reveal any significant effects.

5.5 Discussion

This chapter reported the results from a picture naming and a language switching experiment with a relatively homogeneous group of Norwegian-English bilinguals which have not been studied using these paradigms previously. The data were collected over three sessions, the first two with picture naming in one language each time (either English or Norwegian first) and picture naming with language switching in the third and final session. The analyses of the picture naming and switching experiments included the effects of cognate status, naming language, and picture naming session. The analysis of the switching task additionally included the effect of trial type (switch or stay trial). Two factors derived from self-reported variables relating to bilingual profile, *General L2 proficiency* and *L2 accent and interest*, as well as performance on a flanker task were added as predictors to models of reaction time and accuracy on the word production experiments to assess the effect of individual differences. The discussion will first consider how the current study compares to previous studies of bilingual word production, as well as the effect of language order in the picture naming sessions, before considering the results relating to the predictors.

Several of the results are in line with the findings from the picture naming and language switching studies described earlier in this chapter, both replicating and extending previously observed patterns of results. In what follows I will first discuss these findings before turning to the effects relating to the individual differences predictors. The current study replicated the finding that L1 naming was faster than L2 naming in simple picture naming, and reversed in the switching task, possibly indexing global inhibition of the L1 in the latter. There were also significantly fewer errors in L1 than L2 in the picture naming task, while there was no significant difference between the two languages in the switching task. Switch trials were both significantly slower and more error prone than stay trials. Both the picture naming task and the language switching task reported here yielded robust cognate facilitation effects on naming latencies, extending this observation to a new group of bilinguals. There were also significantly fewer errors for cognates compared to noncognates in the picture naming task, however there was no significant main effect of cognate status on errors in the switching task. Cognate facilitation effects suggest that due to the similarity in both form and meaning, cognate representations are activated cross-linguistically, and their representations are more easily accessed than noncognates, which leads to cognates being named more quickly.

During language switching both languages are activated and compete for selection, and global inhibition suggests the involvement of a mechanism that makes the L1 representations less available to ease the retrieval of L2 representations. As discussed in Section 5.2.2, this mechanism could possibly involve global inhibition or a raised activation threshold for L1 selection. Crucially, for cognates, the activation will be higher than for noncognates, regardless of the language, and this is what the current study found. While there was no interaction between cognate status and naming language in the simple picture naming task, this interaction was significant in the more challenging switching task. While cognates were found to significantly facilitate production in both languages, the benefit was stronger for L1 Norwegian. This pattern is reflected numerically in the error rates. Furthermore, a nested model found that the difference in reaction times between the two languages was significant for noncognate words, but not for cognate words. This pattern is consistent with cognate facilitation ameliorating the effects of L1 inhibition during switching.

The participants in this study named pictures in three separate sessions spanning a maximum of five days. In the first two sessions they completed the simple picture naming task with L1 Norwegian in the first session and L2 English in the second session or vice versa. Language switching always took place in the third and final session. This resembles a blocked switching design spanning several days. To my knowledge, it has not been investigated how naming behaviour might be affected by preceding naming sessions on separate days, and how it might interact with naming language, cognate effects, and switching costs. To assess the influence of language order in the simple picture naming sessions, picture naming session (Session 1 or Session 2) was included in the statistical models of both the picture naming data and the language switching data. There was no significant main effect of session on reaction times in picture naming or language switching, but across experiments, session interacted with cognate status, naming language, and trial type. However, the main effect of session was significant for accuracy rates in both the picture naming task and the language switching task. That is, fewer errors were produced in the second picture naming session compared to the first, and similarly for the switching task, fewer errors were made on trials where participants spoke in the language used in the second picture naming session compared to the language used in the first picture naming session.

In the picture naming experiment, there was a significant three-way interaction between cognate status, naming language, and session. In the switching task there was a significant interaction between cognate status, trial type and session. These will be discussed in turn. There was no significant interaction between cognate status and naming language on the picture naming task. However, as mentioned above, there was a significant three-way interaction between cognate status, naming language and session. Participants who named pictures in Norwegian in the first picture naming session were faster in naming both cognates and noncognates compared to those who named pictures in English in the first session. For naming in the second session the pattern is reversed. This could be interpreted in terms of long-term inhibition. While the cognate benefit remains robust, the L1 benefit is removed when the participants name the pictures in English before naming them in Norwegian. The duration between naming the pictures in each language varied from a night to a maximum of two intervening days and this suggests that L1 inhibition could persist for several days. L1 noncognates are the most negatively affected by being named in the second session (reaction times are on average 74.61 ms longer than in the first session), while the largest benefit is found for L2 cognates in the second session (reaction times are on average 64.10 ms faster than in the first session).

Language order in the picture naming sessions was also involved in significant two-way interactions with both cognate status and naming language and a closer look at these two can inform the interpretation of the three-way interaction.

Cognates were named faster than noncognates in both sessions, but the benefit was larger in the second session (M = 82 ms) compared to the first session (M = 49 ms). Cognates were named more quickly in the second session (M = 756 ms) compared to the first session (M = 776 ms), while noncognates were named more slowly in the second session (M = 838 ms) compared to the first session (M = 825 ms). For cognates, recent naming of the word in a different language facilitated production, due to form similarity, while the recent activation of noncognate translation equivalents inhibited production. There was also a significant interaction between cognate status and session for the error rates, both cognates and noncognates were produced with fewer errors in the second session compared to the first session, however the difference between accuracy on cognate compared to noncognate trials was larger in the second session. That is, similarly to the reaction time data, there was a cognate benefit in both sessions, but the effect was larger in the second picture naming session.

Session also interacted with naming language. While there was an overall L1 benefit in naming, in line with previous studies, and when the pictures were named in L1 during the first experimental session, this effect was reversed in the second naming session, where pictures were named more quickly in L2 than in L1. This could be interpreted in terms of long-term inhibition of the L1 and suggests that L1 inhibition could persist for several days, as mentioned above. When naming pictures in the first session, participants perform as would be expected, with shorter latencies for the more dominant L1. However, when the pictures had been named in L2 in the first picture naming session the participant likely had to inhibit their L1, and in addition the L2 names had been more recently activated leading to slower retrieval of L1 names. These findings support the interpretation of the significant three-way interaction between cognate status, naming language and session.

The effect of language order in the picture naming sessions also seems to influence performance on the switching task, where there was a significant interaction between picture naming session, cognate status, and trial type. For the language that had been named in the session directly preceding the switching task (i.e., in the second session), the interaction seems to reflect a recency effect. While there was cognate facilitation in all conditions, submodels revealed that the three-way interaction is driven by a significant interaction between noncognates and session on stay trials. While average naming latencies were numerically faster for all conditions in the language that was used in the second picture naming session, the largest reduction in reaction times was for noncognates on stay trials. None of the other submodel interactions were significant. For the error rates there was a marginal interaction between cognate status and picture naming session (p = .055). As for the reaction times, there was general cognate facilitation (fewer errors on cognate trials compared no noncognate trials) and fewer errors on trials produced in the language that was used in the second picture naming session compared to the first picture naming session. The largest reduction in error rates is found for noncognate trials produced in the language produced in the second picture naming session.

Since naming language is not involved in the interactions reported above, this seems to reflect a language non-specific recency benefit on stay trials in the most recently used language. This suggests that cognates behave similarly in the switching task irrespective of what language they were first named in, while noncognates show a small recency effect, i.e., a benefit on stay trials in the most recently used language. The effect of session in the simple picture naming task is consistent with long term L1 inhibition. In contrast, I have interpreted the effect of session in the switching task to language recency. However, there is evidence that lexical repetition effects can persist over long periods of time (Wheeldon & Monsell, 1992). It remains possible therefore that the effect of session may primarily be located at the level of lexical activation rather than of language activation. A further study manipulating language change without lexical repetition would be required to test this possibility.

The language switching task included three theoretically relevant factors that have been investigated in previous studies, that is cognate status, naming language and trial type (stay or switch). The current study replicated several previously reported findings with a new group of bilinguals. To summarise, there was a significant main effect of trial type, where stay trials were produced with shorter latencies than switch trials, and naming language, where the L2 latencies were shorter than L1 latencies. Previous studies have suggested that switch costs (the difference between reaction times from stay and switch trials) are asymmetric (i.e., longer latencies when switching from L2 into L1 than vice versa) for unbalanced bilinguals, while switch costs are often symmetric for more balanced bilinguals (e.g., Costa & Santesteban, 2004; Meuter & Allport, 1999; but see Christoffels et al., 2007). There was no interaction between trial type and naming language in the current study, and therefore, as hypothesised, switch costs appear to be symmetric in the current population. Finally, there was a significant main effect of cognate status, where pictures with cognate names were produced more quickly than pictures with noncognate names in all conditions.

While some patterns of results reoccur in the switching literature, such as general cognate facilitation on reaction times (e.g., Christoffels et al., 2007; Santesteban & Costa, 2016; Verhoef et al., 2009), findings are more mixed regarding the magnitude and direction of interactions between cognate status, naming language and trial type. This in turn has led to different theories and explanations of the effects and cognitive mechanisms involved in language switching with cognates. Comparing results across these studies is not straightforward, due to different design decisions, such as manipulating cognate status within (Christoffels et al., 2007) or between experimental blocks (Declerck et al., 2012), manipulating cognate status preceding a critical trial (Santesteban & Costa, 2016) or just focusing on the cognate status of the critical trials (Christoffels et al., 2007), repetition of critical stimuli (C. Li & Gollan, 2018) or only one presentation of each (Broersma et al., 2016), and different lengths of preparation time between a language cue and the presentation of a picture to be named, for instance no preparation time (Broersma et al., 2016), or manipulation of preparation time within an experiment (Verhoef et al., 2009). While it might not be possible to separate these effects and their explanations, or understand how they interact at this point, some of the papers summarised in Section 5.2.3 will be considered in more detail below and compared to findings in the current study.

The analysis of the picture naming and language switching tasks in the current study has focused on reaction times and accuracy. As mentioned above switch costs in the current study were symmetrical, i.e., not significantly different, when switching from L1 to L2 and vice versa. While cognates generally facilitate reaction times, another issue to consider is whether cognates facilitation extends to switch costs. Christoffels et al. (2007) report a study with both simple picture naming and language switching, as in the current study. In both studies, significant main effects in the simple picture naming task and language switching went in the same direction, despite differences in participants' bilingual profile, the number of participants, and the experimental design. There were also some key similarities in the design of the studies, such as the same number of stimuli and repetitions, as well as the stimulus picture and the language cue being presented simultaneously. In addition to faster naming in L1 than L2 in simple picture naming and the reversed pattern in language switching, they found that the cognate facilitation effect was larger for L2 than L1 in simple picture naming and vice versa in the switching task. As discussed above, a similar result was found for the switching experiment in the current study, that is, a stronger cognate facilitation effect for the L1 compared to the L2 in the language switching task. These findings provide compelling evidence of cross-linguistic phonological activation and that cognates can facilitate production in both L1 and L2. There is however one key difference in the results. While cognates were uniformly helpful in the current study, including reduced switch costs for cognates compared to noncognates, switch costs in Christoffels et al. (2007) were higher for cognates compared to noncognates, i.e., suggesting cognate inhibition, as pointed out by C. Li and Gollan (2018).

One key methodological difference between the switching study in Christoffels et al. (2007) and the current study is that cognate status was mixed within blocks in the former while cognate status was blocked in the latter. Previous studies, with preparation time between the language que and stimuli to be named, have also found that switch costs were facilitated by cognates when cognate status was blocked (Declerck et al., 2012), but not when cognate status was mixed within blocks (Verhoef et al., 2009). This methodological difference was addressed in three experiments by C. Li and Gollan (2018). In the first experiment, where cognate status was blocked, they found significant main effects of trial type and cognate status, as well as a marginal effect of naming language, all of which went in the same direction as described above. In addition, they found smaller switch costs for cognates compared to noncognates in the dominant language, but not the nondominant language.

The second experiment in C. Li and Gollan (2018) was similar to the first, but cognate status was mixed within blocks. The main effects were similar, but when cognate and noncognate stimuli were mixed they did not find evidence of cognates modulating switch costs, apart from in an analysis only including the first presentation of stimuli. Comparing the first and second experiment they also found that while repetition did not affect reaction times in the first experiment, nor noncognate reaction times in the second experiment, cognates were produced more slowly with repetition in the second experiment. In the second experiment there were more items between each repetition of a given stimulus than in the first experiment. Therefore, a third experiment was conducted, with the same design but fewer items than in the second experiment, in order to separate the effects of lag between repetitions from the effect of words with different cognate status being presented in separate or mixed blocks. Again, they found similar main effects of trial type and naming language, but cognate naming was not significantly faster than noncognate naming in this experiment. In this final experiment, switch costs were significantly larger for cognates than noncognates, indicating cognate inhibition. This shows that the observed interactions between naming language, cognate status and trial type depend not only on whether experimental blocks are separated by cognate status or not, but also the number of items and repetitions. In the current experiment, each item was repeated four times (once in every condition: L1 switch, L1 stay, L2 switch, L2 stay) and it is therefore possible that cognate facilitation effects might have disappeared with more repetition. However, it remains unclear how repetition would influence processing in a switching experiment where items are blocked by cognate status.

In the current study, the results do not show any clear effects of naming language interacting with trial type in the switching task. There is a marginal interaction between the *L2 accent and interest* predictor, trial type and naming language which will be discussed later. Considering the clear effects of naming language, trial type and cognate status overall in the current study, a closer look at the interaction between these three factors may be informative, even though it did not reach significance (p = .081). Cognate trials were faster than noncognates in all conditions, but on average the benefit was smallest for L2 stay trials (M = 61.96 ms) compared to the other conditions (L1 stay M = 100.78; L1 switch M = 102.29; L2 switch M = 104.04). Now turning to the switch costs, they were similar in size for L1 cognates (M = 56.41 ms) and L1 noncognates (M = 57.92 ms), but L2 switch costs were smaller for cognates (M = 43.50 ms) compared to noncognates (M = 85.7 ms), i.e., the cognate facilitation of switching costs was only found for the non-dominant language. In contrast, the first experiment in C. Li and Gollan (2018) found cognate facilitation on switch trials for the dominant language only. Their Spanish-English bilingual participants were English dominant, but the majority were L1 speakers of Spanish or had learned both languages simultaneously. In contrast, the participants in the current study acquired their dominant language Norwegian before starting to acquire English and live in a mostly Norwegian speaking environment. This could suggest an influence of both variation in language acquisition and the current language dominance of the speaker.

Interestingly, the pattern of switch costs in the significant three-way interaction between cognate, trial type and picture naming session in the current study (discussed above) is quite similar to the pattern found in the non-significant three-way interaction between cognate status, trial type and naming language. That is, switch costs are relatively similar for cognates (M = 55.51 ms) and noncognates (M = 59.84 ms) when produced in the language named in the first picture naming session, while for the most recently named language (i.e., named in the second picture naming session) switch costs are smaller for cognates (M = 44.59 ms) compared to noncognates (M = 83.78 ms). This means that, on average and regardless of language order in the picture naming sessions, the effects of cognate status on switch costs are comparable for L1 Norwegian and the language used in first picture naming session (where half of the participants named pictures in L1 Norwegian, and half named pictures in L2 English). Similarly, switch costs found for cognates and noncognates in L2 English show a similar pattern to the language used in the second picture naming session.

Both reaction times in the simple picture naming sessions themselves and the switching task are affected by language order in the two single-language picture naming sessions. This effect was not observed in a recent experiment conducted in our lab where single language picture naming and language switching experiments were conducted on the same day (M. Albrecht, personal communication, 2021). This could suggest that sleep consolidation of the association between pictures and their labels on the first day (in either L1 or L2) leads to these picture-name associations being processed similarly to an L1, while picture-name associations made on the second day (in either L2 or L1) behave more like an L2. In the discussion above, the effect of session in the simple picture naming task was interpreted in terms of long term L1 inhibition while the effect of session in the switching task seemed to reflect a language non-specific recency benefit on stay trials in the most recently used language. It is possible that the picture-word associations formed for the language used in the first picture naming session is processed similarly to an L1 and that the language used in the second picture naming session is processed similarly to an L2. In this case, this effect could also be interpreted in terms of inhibition of the initially formed associations. That is, inhibition is applied to the language named in the first picture naming session, regardless of it being L1 or L2. There are robust effects of naming language overall in the current study, however this represents a potential confound that should be considered in the design of future studies. Future studies could also work to separate these effects.

Alternative explanations of switching effects have also been put forward. For instance, Costa and colleagues (Costa et al., 2006; Costa & Santesteban, 2004) argue that reactive inhibition might be involved in bilingual language control when proficiency in one language is much weaker than the other, but that more proficient bilinguals have developed a mechanism that is applied to the whole language, such as a language-specific activation threshold for selection. Christoffels et al. (2007), discussed in detail above, recorded event-related brain potentials (ERPs) in addition to behavioural measures. They argue that their results support a role for proactive/global inhibition in language control, but to a lesser extent support inhibition of lexical items. Language control might therefore be achieved by inhibition combined with other mechanisms, such as language-specific activation of lexical items. However, effects of cognate status also suggest the presence of activation that is not language specific.

In summary, previous studies suggest that there might be several cognitive processes that interact to produce the target output during language switching. These may also influence whether cognate effects emerge during language switching. The studies discussed so far have all differed methodologically, however the analyses have focused on how measures from specific trials are affected by cognate status, naming language and trial type. Santesteban and Costa (2016) compared two groups of bilinguals, unbalanced vs highly proficient, and investigated both whether language switching was affected by the cognate status of a given target trial and the cognate status of the preceding trial. They found that cognates were named faster than noncognates in general, but there was no evidence for cognate facilitation of switch costs in either participant group. Compared to the current study there are some key methodological differences, including fewer items, more repetitions of each target word, and an interval between the language cue and the target. Considering the discussion above, it seems likely that limiting repetitions and preparation time may be important for cognate effects to be observed.

In Broersma et al. (2016) target cognates and controls were only presented once, however these were embedded within filler trials. Participants were three groups of Welsh and English speaking bilinguals who differed in language dominance. Switching only occurred on filler trials. Therefore, the analysis of switching behaviour focused on the effects of cognate status in the preceding trial, while the effect of cognate status and naming language were assessed in the critical trials. As described in Section 5.2.3 they found cognate inhibition in Welsh and no cognate effect in English for the English dominant participants, while cognates generally facilitated production for both the Welsh dominant participants and the more balanced group. In addition, they found evidence of cognate inhibition of the following trial in both languages for all three groups. That is, filler trials named after cognate critical trials were slower than those named after noncognate critical trials. These results are interpreted in terms of two parallel processes, competition for selection at lexical-semantic level and facilitation at the word form level. Therefore, depending on the relative strength of activation at these levels of processing cognate facilitation, inhibition, or no cognate effect may be observed.

C. Li and Gollan (2018), as discussed above, found that with repetition of stimuli cognate facilitation of switch trials could be reduced or reversed. They argue that this could reflect a problem discriminating phonologically similar segments (in cognates) when they have been named several times in different languages. They argue that while there is facilitation at the phonological level, feedback to the lexical level could eventually lead to more competition in lexical selection. The question is then, what determines the level of activation at the different levels of processing. This discussion has focused on methodological differences in experimental procedures, but the studies mentioned also test participants with different language backgrounds and proficiency levels. Broersma et al. (2016) hypothesise that the observed cognate inhibition in Welsh for the English dominant participants could be related to the fact that the Welsh-English cognates originally were borrowed from English. Clearly the phonological similarity between cognates in two languages will depend on the languages involved, the direction of the loan, and the proficiency of the speaker, and this should also be considered more carefully in future studies.

As there are mixed findings in the literature, there are still many open questions regarding how languages are controlled during bilingual word production. What is clear, however, is that language control demands and the language production process are affected by several factors, such as the proficiency of the speaker, preparation time, repetition, and phonological similarity. Now the discussion turns to the individual difference predictors employed in my study to address the two research questions asked at the beginning of this chapter. In the current study, where the focus is on cognate effects in speech production, both general language proficiency and individual differences in the domains of phonology and accent are of particular interest. While cognates share meaning and form between two languages, the degree of form overlap or similarity in produced speech and phonological representations is subject to individual differences. These differences might in turn impact the cognate effect. Two factors obtained from the factor analysis in Chapter 4 were added as predictors in the models of the simple picture naming and language switching data. In addition, the participants' performance on a version of the flanker task was included as a non-linguistic measure of attentional control. This section will start by discussing the results from the analysis with the flanker task predictor, before discussing the results relating to the bilingual profile predictors, *General L2 proficiency* and *L2 accent and interest*.

The first research question in this chapter asked: How do individual differences in non-linguistic attentional control interact with cognate effects in language production and switching? The difference in latencies between congruent and incongruent trials on the flanker task served as a non-linguistic measure of selective attention and control. In the analysis of latencies from the simple picture naming task, the main effect was not significant, but there was an interaction between cognate status and the flanker score. Smaller RT differences between congruent and incongruent trials in the flanker task were associated with faster responses for both cognates and noncognates, but the effect was larger for noncognates compared to cognates. This could suggest that there is cross-language activation in the simple picture naming task, even though items are only produced in one of their languages. Producing noncognate words, with different phonological forms in the two languages, likely requires more control than producing cognate words.

The interaction with flanker and cognate did not survive in the more complex switching task. Here there was a main effect of the flanker score, again showing faster naming with smaller RT differences between congruent and incongruent trials on the flanker task. These results suggest that there is a general production benefit of better non-linguistic attentional control in speech production. This view is supported by the accuracy analysis of both tasks, where there were significantly fewer errors with smaller RT differences between congruent and incongruent trials, and the fact that flanker scores did not interact with trial type on the switching task. The flanker score did not interact with naming language or picture naming session in either of the tasks, nor cognate status and trial type in the switching task, consistent with the involvement of additional control mechanisms in linguistic processing.

The second research question addressed in this chapter is: Do aspects of bilingual profile modulate naming and switching behaviour? Two of the factors extracted from the questionnaire data (detailed in Section 4.4.2), *General L2 proficiency* and *L2 accent and interest*, were added to models of latencies and error rates in the simple picture naming task and language switching task to inform this question. There were no significant effects of the *General L2 proficiency* predictor in the analysis of reaction times from the picture naming task nor the switching task. This could simply reflect that the participants in this study have relatively similar levels of proficiency, at least as captured by the questionnaire, and therefore the variation in L2 proficiency observed within the group is not large enough to account for differences in naming behaviour. There was a marginal interaction between cognate status, naming language and *General L2 proficiency* in the switching task. For all conditions naming latencies were reduced with higher levels of the L2 proficiency predictor, but while the effect was similar in size for L2 cognates, L2 noncognates, and L1 noncognates, the effect appears to be slightly larger for L1 cognates.

In the analysis of the error rates there was no effect of the *General L2 proficiency* predictor in the simple picture naming task, but in the switching task there was a significant 4-way interaction between cognate status, trial type, naming language and the predictor. In both languages there were more errors on switch trials than on stay trials. The interaction seemed to be driven by a significant 3-way interaction between cognate status, trial type and the *General L2 proficiency* predictor in L1 Norwegian. Further analysis did not reveal any significant effects, but Figure 19 suggests that higher scores of the *General L2 proficiency* predictor generally associate with lower error rates, but not for L1 noncognates. This was partially reflected in the naming latencies, where the marginal interaction between cognate status, naming language and the predictor suggested that higher scores of the predictor generally reduced latencies. Since trial type was not involved in this interaction, no speed-accuracy trade-off when switching into L1 noncognates was detected in the analysis. These effects are not very strong and difficult to interpret. Overall, L2 proficiency, as measured by the predictor, does not appear to modulate naming and switching behaviour in this group of bilinguals.

The *L2 accent and interest* predictor main effect was significant in the simple picture naming task, reflecting that naming generally was faster with higher values of the predictor. The analysis of the significant three-way interaction with cognate status and naming language showed that L2 cognates were named more quickly as the value of the predictor increased, while L2 noncognates showed a small effect in the opposite direction. In the L1, higher predictor values lead to faster naming of both cognates and noncognates, an effect that was marginally larger for noncognates. Therefore, the L1 difference between cognate and noncognate latencies decreases with higher values of the predictor, and in L2 this difference increases with higher values of the predictor. The analysis of error rates on the picture naming task showed a significant interaction between naming language and the predictor however the nature of this interaction is unclear as the effect was not significant in either language-specific submodel.

As discussed in Chapter 4, obtaining self-ratings in the domains of phonology and accent is not straightforward. The term was intended as a broad label to cover variables assessing attitudes and proficiency related both to phonology and accent, as individual differences in these domains were hypothesised to relate to cognate effects and the speech production process. The *L2 accent and interest* factor is associated with higher scores on questions relating to interest in, and awareness of, L2 accent and pronunciation, and two language general variables ("paying attention to other people's pronunciation" and "ability to imitate accents"). As such, this is not a clear proficiency or ability factor. It is possible that participants with high scores on this factor are more focused on pronouncing L2 correctly, possibly slowing speech production, which could explain why the only negative relationship is found for L2 noncognates. This does not, however, explain why naming in the other conditions is faster with higher values of the predictor. As mentioned, the *L2 accent and interest* factor contains positive loadings of two language general variables. In addition, the analysis in Chapter 4 found that the factor correlated with L2 spelling and L1 elision scores, but none of the other test scores. It is possible that higher values on this factor partially reflect more language general aspects that are associated with faster production. One can also speculate on whether there is a "second session" effect that is less problematic for cognates, which could lead to a larger difference between cognate and noncognates, however this was not indicated by the analysis.

In the switching task, the factor was dropped from the model of accuracy, and there were no significant results involving the *L2 accent and interest* predictor in the analysis of the reaction time data. There was however a marginal (p = .055) 3-way interaction for the naming latencies between the predictor, trial type and naming language. While the main effect of the predictor was not significant in the analysis, the general pattern showed shorter naming latencies with higher values of the predictor, as in the picture naming task. The analysis of the three-way interaction suggested that the predictor had a larger effect on L1 Norwegian, especially on stay trials, compared to stay and switch trials in L2 English. There appears to be a greater speeding of responses with higher levels of the factor in stay trials compared to switch trials in L1. In L1, the lowest factor scores are associated with slow naming overall and little to no switch costs. It is not clear why this effect is observed for the L1. One possibility is a build-up of L1 inhibition for less proficient speakers that persists across stay and switch trials.

In general, the effects of the predictors are not very strong. Considering the pattern of residuals for extremes of the latency distribution (Appendix J), the results, and especially the interpretations presented above, must be treated with caution. To summarise, the analyses of latencies from the picture naming task and the switching task found no significant effects of the *General L2 proficiency* predictor. The marginal interaction between cognate status, naming language and the predictor in the switching task suggested that naming latencies were reduced with higher levels of the predictor in all conditions, but that the effect was slightly larger for L1 cognates. There were significant effects of the *L2 accent and interest* predictor in the picture naming task, but not in the switching task. In the picture naming task naming was faster with higher values of the predictor overall, but the three-way interaction showed that this was not the case for L2 noncognates. In the switching task, a marginal interaction between the predictor, trial type and naming language also suggested that higher values of the predictor were related to shorter naming latencies and that the effect was larger for L1, especially on stay trials.

The aspects of bilingual profile considered in the current study, *General L2 proficiency* and *L2 accent and interest* were to a limited extent found to modulate naming and switching behaviour. The *General L2 proficiency* factor was not significant in the analysis of the picture naming task nor the switching task. In this group of relatively proficient bilinguals, this finding is in line with previous studies. Of particular interest in the current study was to investigate whether individual differences in the domains of phonology and accent would modulate naming and switching behaviour, especially cognate effects. The *L2 accent and interest* predictor significantly modulated behaviour in the simple picture naming task. Naming latencies were generally faster with higher values of the predictor, but a three-way interaction with cognate status and naming language indicated that this was not the case for L2 noncognates. Several possible explanations were put forward in the discussion, but it remains unclear why the results are different for L2 noncognates. In the more challenging switching task, there were two marginal interactions, one with each of the predictors. Both suggested that naming was faster with higher levels of the predictor in general, and although they involve different variables, both imply that higher values of the predictors are more beneficial for L1 latencies than L2 latencies. That is, the marginal interaction between the *General L2 proficiency* predictor, cognate status and naming language suggested that the benefit was slightly larger for L1 cognates and the marginal interaction between the *L2 accent and interest* predictor, trial type and naming language suggested that the benefit was larger in L1, especially on stay trials.

The limited effects observed for the *L2 accent and interest* predictor and the *General L2 proficiency* are in the direction of faster production with higher values of the predictors, and these seem to benefit L1 production the most. This tendency was found in both the picture naming and the switching task. In the language switching task the predictors were involved in one marginal three-way interaction each, both involving naming language. The *L2 accent and interest* predictor additionally interacted with trial type, and the *General L2 proficiency* predictor with cognate status. As discussed above, previous studies have suggested several processes that may interact to modulate latencies and cognate effects in picture naming, and particularly with language switching. The observed effects in the current study are marginal and complex, it is therefore challenging to separate these effects and there is not enough experimental support to make any strong claims.

As discussed in this chapter, several findings in the bilingual language production literature have shown an effect of general language proficiency in bilinguals' languages. In the current study, with relatively homogenous, proficient, L1 dominant bilinguals the effect of proficiency differences was investigated by a measure of general L2 proficiency. This was not found to modulate naming and switching behaviour. Cognate effects suggest that there is cross-language activation at the level of phonology. It is therefore possible that individual differences in the domains of phonology and accent specifically could modulate naming and switching behaviour, including cognate effects, even though individual differences in general L2 proficiency did not in this group of participants. This was the case in the simple picture naming task, where higher values of the *L2 accent and interest* predictor were associated with faster naming of cognates in both L1 and L2, in addition to L1 noncognates. The results from the switching task on the other hand did not show any evidence of the *L2 accent and interest* predictor modulating cognate effects.

Previous studies have found both cognate inhibition and facilitation in switching tasks, suggesting that the phonological similarity of cognates might not always be beneficial, especially in language switching. As discussed above, a number of possible explanations have been put forward, including facilitation at the phonological level and competition for selection or problems discriminating phonological feedback at lexical-semantic level. In the more complex switching task, the bilingual profile predictors were involved in two marginal interactions, and some preliminary interpretations were put forward. The *L2 accent and interest* predictor employed in this chapter was based on self-reported attitudes and abilities. A future study with a more targeted measure of phonological abilities and more power might be able to expand on some of the weak effects observed in the current study.

5.5.1 Conclusion

This chapter reports an investigation into the effects of cognate status and individual differences in bilingual profile on spoken word production and language control. In three experimental sessions, spanning a maximum of five days, participants completed two simple picture naming tasks (once in L1 and once in L2) and a language switching task. In both the picture naming and the language switching task cognate facilitation effects were robust. Like in previous studies, a reversed dominance was observed in the language switching task (faster naming in L2 than in L1) indicating L1 inhibition. Furthermore, the inclusion of picture naming session in the model of the switching task data (i.e., whether pictures were named in a given language in the first or second session), suggests that L1 inhibition could persist for days. However, further research is needed to separate the effects of inhibition and repetition priming.

Two research questions were posed at the beginning of this chapter. The first asking how individual differences in non-linguistic attentional control interact with cognate effects in language production and switching. Better attentional control (as measured by the flanker task) was associated with faster naming in both languages and on both tasks, but in the simple picture naming task the effect was only significant in interaction with cognate status. The flanker task score did not interact with any of the other variables, suggesting the involvement of additional control mechanisms in linguistic processing.

The second question asked whether aspects of bilingual profile modulate naming and switching behaviour. There were no significant results involving the *General L2 proficiency* predictor, but the *L2 accent and interest* predictor significantly modulated latencies in the picture naming task. Higher values of the *L2 accent and interest* predictor were associated with faster naming in general, but a three-way interaction suggested this was not the case for L2 noncognates. In the switching task each of the predictors were involved in a marginal interaction. The limited effects observed are generally in the direction of faster production with higher values of the predictors, however considering the marginal and complex nature of the interactions, there is no strong evidence of the measures modulating naming and switching behaviour, perhaps apart from the *L2 accent and interest* predictor in the simple picture naming task.

In the current analysis predictors were limited to two theoretically motivated factors, *L2 accent and interest* and *General L2 proficiency*, and some aspects of the results were not readily interpretable. It is possible that other measures collected in this study can help clarify the role of the L2 accent and interest predictor and the potential influence of individual differences in the domains of phonology and accent. This question will be revisited in Chapter 6, where measures of articulatory divergence between L1 and L2 productions compared to language test scores and specific self-ratings from Chapter 4. The articulatory divergence measures are also entered into models of the picture naming and switching data.

6 Individual differences in articulation

6.1 Introduction

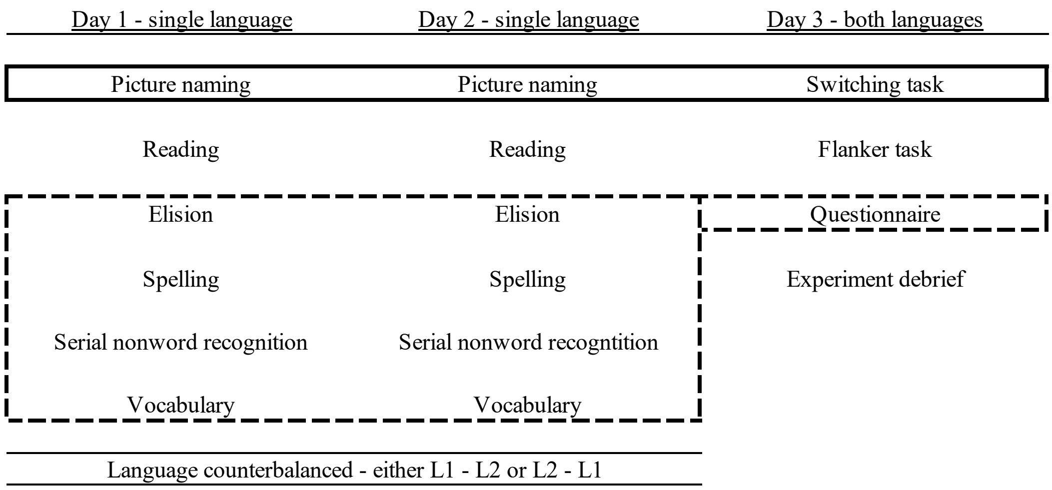
In previous chapters the focus has been looking at how self-ratings in the domains of phonology and accent relate to other aspects of bilingual profile, objective language tests, bilingual word production, and language control. Bilinguals show considerable variation in the degree to which they adapt their L1 articulation when speaking an L2. In this chapter, the focus turns to an acoustic analysis of vowel production, and how differences between L1 and L2 productions relate to the measures discussed in the previous chapters. Previous studies have reported on how differences in bilingual language profile and language proficiency may modulate bilingual language production (Bonfieni et al., 2019), including language switching (Prior & Gollan, 2011) and cognate effects (Costa et al., 2000). Both cognate facilitation (Declerck et al., 2012) and inhibition (C. Li & Gollan, 2018) effects have been reported in switching tasks, suggesting that the phonological similarity of cognates might not always be beneficial. One explanation for these findings is that while there is cognate facilitation at the phonological level, there could be competition for selection or problems discriminating phonological feedback at lexical-semantic level (Broersma et al., 2016; C. Li & Gollan, 2018). Depending on the combined activation at each of these levels the result could therefore be either facilitation or inhibition. The degree of phonological similarity between cognates does not only depend on the languages involved, but also individual differences in representations of sound structure.

The Speech Learning Model (SLM; Flege, 1995, 2007), and the recently revised version (SLM-r; Flege & Bohn, 2021), posits that bilingual representations of sound structure exist in a “common phonological space”. Herein, L1 and L2 representations can mutually influence each other, both in perception and production. The SLM(-r) predicts that individuals with more precise L1 categories are more likely to perceive a difference between similar L1 and L2 speech sounds, and in turn establish new representations for L2 speech sounds. As previously discussed, assessing proficiency and individual differences in the domains of phonology and accent is challenging. In the current chapter, individual differences in articulation are assessed by measuring the articulatory divergence between two pairs of similar, but not identical, L1 and L2 speech sounds. It is hypothesised that a larger acoustic difference between similar L1 and L2 speech sounds (i.e., more divergent articulation) indicates more precise L1 and L2 speech sound categories compared to smaller acoustic differences.

Two measures of articulatory divergence, the Pillai score (first used to assess vowel overlap by Hay et al., 2006) and Bhattacharyya's Affinity (BA; Bhattacharyya, 1943), are calculated for two pairs of similar L1 and L2 vowels (L2 /uː/ - L1 /ʉː/ and L2 /ʌ/ - L1 /œ/). Speech samples were collected during the picture naming task and the language switching task (marked by the black frame in Figure 20). An exploratory approach is taken to investigate whether these measures relate to the self-reported and tested proficiency measures, as well as bilingual profile factors, obtained in Chapter 4 (marked by stapled frames in Figure 20). Finally, the divergence measures were entered as predictors in models of the picture naming and the language switching latencies.

This chapter starts with a theoretical overview, focusing on the representation of sound structure. Starting with the acquisition of first language sound structure, then turning to the representation of sound structure in L1 and L2 processing models, before looking at effects of cognate status and language switching on articulation. The method section includes a discussion of methodological considerations and methods for measuring vowel formants and articulatory divergence, before comparing English and Norwegian vowel inventories and describing the vowels of interest in the current study. Then the development of a forced aligner is detailed before formant measures from the aligned vowels are evaluated and refined. In the result section the final divergence measures are presented first. The analyses comparing the divergence measures to self-reported and tested proficiency measures, bilingual profile factors and latencies from picture naming and language switching tasks are presented in separate sections, followed by a general discussion of the results.

Figure 20
Experiment overview Chapter 6



Note. Speech samples were collected during experiments in the black frame. Experiments in stapled frames provide measures of language proficiency and bilingual profile.

The research questions addressed in this chapter concern the relationships between individual differences in L1-L2 speech divergence, bilingual profile, and word production. Specifically:

- Does L1-L2 speech divergence relate to self-rated pronunciation proficiency?

- Does L1-L2 speech divergence relate to objective language test scores?

- How do L1-L2 speech divergence measures relate to aspects of bilingual profile?

- Does the degree of articulatory divergence impact language production behaviours in picture naming and language switching?

The speech divergence measures did not relate to self-ratings of accent proficiency. Some of the divergence measures correlated with L1 serial nonword recognition (SNWR) and L2 spelling scores. This could suggest that there is a link between the divergence measures and tested proficiency in domains related to phonology, however no effects were observed for the elision scores in either language, L2 SNWR scores, or L1 spelling scores. Five bilingual profile factors were related to the divergence measures, but only for one of the vowel pairs. The discussion considered why different results were observed depending on the vowel pair. One possibility is that individual differences in the production of the more acoustically similar vowel pair (L2 /uː/ - L1 /ʉː/) are more related to specific phonological abilities, while more divergent productions of the other vowel pair (L2 /ʌ/ - L1 /œ/) are associated with more L2 exposure. There is, however, only partial support for these interpretations. Finally, there were limited effects of articulatory divergence on language production. The observed effects were in the direction of faster naming with more divergent production, apart from for L1 noncognates in the picture naming task. One clear observation is that different effects were found depending on the divergence measure used (Pillai score or BA) and which vowel pair was used to quantify individual differences in articulation.

6.2 Theoretical overview

6.2.1 Representing sound structure in L1

First language acquisition typically starts with perception, and evidence suggests speech learning, or sound pattern acquisition, starts before birth (e.g., Hepper et al., 1993). During the first year, sensitivity to language specific sound patterns increases (Kuhl et al., 2006), while the sensitivity to variations that are specific to other languages decreases (Werker & Tees, 1984). One view is that this enables the infant to recognise critical speech segments and words in the speech signal, and they start to produce segments and sequences of segments. Meaning is mapped onto units of speech sounds, and these are combined with increasing complexity. In this way perception forms the building blocks for further language development (for a discussion, see Vihman, 2017). A major challenge for language perception in general, and when acquiring a language, comes from variability in the speech signal. There are many sources of variability. Speech segments are for example affected by their phonological or phonetic context (i.e., coarticulation), higher level characteristics such as speaking rate, and speaker-related characteristics such as age, dialect, and speech style.

Models of speech perception deal with this variation in different ways and make different assumptions about the representation of sound structure. Some approaches build sparse representations, dealing with variation through a process of abstraction and/or normalisation. That is, as the speech signal is being perceived, the speaker identifies the salient properties crucial to identifying speech sounds and ignores irrelevant variation (e.g., Lahiri & Reetz, 2002). In contrast, other models propose that detailed representations of speech are stored in full. In these models, statistical properties of language are the main force behind language representation, for instance in exemplar theory for speech perception (e.g., Pierrehumbert, 2001) where word frequency plays an important role. In this section I will discuss some models that focus on processing of speech sounds in a first language. Although the focus of my research is on language production, many models assume a tight link between perception and production in the representation and processing of speech. Therefore, relevant theories of perception will also be described.

The classic speech production model developed by Levelt (1989, summarised in Section 2.3) describes the production process from conceptualising an idea or a message, to the formulation of a phonological and articulatory plan. The final stages of speech production are detailed in Levelt and Wheeldon (1994). After word form retrieval, information about syllables and accent structure is added to form a phonological word. In this model, articulation is determined on a syllable-by-syllable basis. Speakers can access abstract, overlearned gestural scores which specify which articulatory movements (or tasks) are necessary to produce a given syllable structure. The gestural scores are then passed to an articulatory network which controls and monitors the final articulation of speech. They adopt the framework from Browman and Goldstein (1990) where gestural scores are specified in five tiers corresponding to five articulators (the glottal system, the velar system, tongue body, tongue tip, and lips). Within this model phonetic segments only exist as a part of the syllable. The gestural scores contain abstract information about which articulators are involved in the production of a given syllable. The specific motoric movements required to produce the final articulation, controlled by the articulatory network, on the other hand are subject to variation.

The starting point in speech perception is the variable acoustic output that is the result of the speech production process. The goal of speech perception models is to account for how meaningful linguistic content is extracted from the acoustic speech signal. As the representations used in production are built on perception, and we monitor the speech we produce, these processes are inextricably bound. Models and theories of speech perception describe different ideas about the representation of sound structure and the type of information extracted from the speech signal. Representations range from sparse to full representations. Speech sounds may be perceived in terms of features or gestures and there are also different views on the extent to which these representations are perceived directly or abstracted further. A few key theories will be considered below.

One view is that the listener perceives phonetic features, which are the specific acoustic consequences of articulation that cue a particular speech sound. For example, such features signalling /d/ might be prevoicing, voice onset time (VOT) and spectral characteristics (e.g., centre of gravity) alluding to the place of articulation. Stevens (2002) describes a three-stage perceptual model where the acoustic signal is interpreted into phonological segments through acoustic landmarks and the use of binary distinctive features. Herein, segments are represented by a bundle of abstract features. In the lexicon, words are represented as a sequence of segments, and there is information about syllable structures and constraints. In perception, the first step is locating acoustic landmarks (e.g., peaks, troughs, and abrupt changes in the signal) and estimating articulator-free features signalling whether the segment is a vowel, glide, or consonant. Then acoustic cues found close to the landmarks are evaluated. Finally, this information is weighted and combined with suprasegmental information to estimate the value of articulator-bound features. These are then matched to the abstract representations stored in the lexicon. In this model there is a total of seven articulators (lips, tongue blade, tongue body, soft palate, pharynx, glottis, and vocal folds), but features are only thought to be specified for the articulators involved in producing a specific segment.

The Featurally Underspecified Lexicon model (Lahiri & Reetz, 2002) takes a minimalist approach to phonemic representation and the processes involved in speech perception. Only the minimal number of features that are necessary to differentiate speech sounds in a language are represented, so that there is no redundancy in the perception system. The same place features are used for vowels and consonants. In addition, some features are not specified, such as coronal for place of articulation. Acoustic characteristics are extracted directly from the speech signal and converted into phonological features. The set of perceived features are compared to abstract representations of word candidates in the lexicon. The activation of candidates is based on how well the input matches the stored feature information. This can account for how listeners deal with variation in the speech signal, for example the tendency for coronals to assimilate in contrast to dorsals and labials (e.g., Roberts et al., 2013).

An alternative view of the units of perception is found in the motor theory of speech perception (Liberman & Mattingly, 1985). Rather than perceiving speech sounds by extracting phonetic or phonological features from the acoustic signal, listeners perceive the *intended* articulatory gestures of the speaker. That is, the movements and configurations of the speech organs involved to produce a given sound. Liberman and Mattingly (1985) propose that speech perception is achieved through a specialised module that detects these gestures. Within this module there is information about the links between acoustic patterns and the neuromuscular processes necessary to produce speech sounds, and it further assumes a strong link between perception and production. However consensus about the detailed nature of the representations, as well as the link between them in perception and production remains to be reached (for a review, see e.g., Skipper et al., 2017).

The accounts of phonological processing described above all assume some level of abstraction. Another way the perceptual system might deal with variable input is found in exemplar theory for speech perception (e.g., Bybee, 2001; Pierrehumbert, 2001, 2003). Herein, speech sound categories are represented as collections of heard instances (memory traces) of a particular category, rather than one normalised abstract representation. The memory traces are organised in relation to each other on a map according to phonetic parameters, where similar instances are grouped together under labels or categories. In speech perception, stimuli are encoded and activate exemplars that are close to each other on the map, activation spreads up to the labels, and the most probable label is selected. One key criticism of exemplar type models is the memory load involved, considering the amount of speech, and therefore also the number of speech sounds, a person will encounter. Pierrehumbert (2001) offers a solution in that the representations are granularised. That is, the model assumes that the strength of each exemplar is affected by the frequency and recency of the memory traces in that area of the cognitive map. Similarly, Bybee (2001) proposes representations somewhere between prototypes and exemplars. These types of models can account for variation by representing several instances of speech sounds, and language change, as the system is affected by the frequency and recency of speech input. The focus now turns to representation of sound structure in L2.

6.2.2 Representation of sound structure in L2

Several of the representations and processes described for L1 speech processes are found in theories and models of L2 speech processing. L2 models are not only concerned with the representations and mechanisms involved. A key difference, compared to L1 speech processing, is of course that the speaker already knows the sounds, rules, and structures associated with their native language, and these may influence the acquisition and use of an L2. The models of L2 speech processing considered here focus less on the exact nature of the processes and representations involved in speech processing, but rather how speakers' languages interact in perceiving and learning the sound system of a new language. The Perceptual Assimilation Model L2 (PAM(-L2); Best & Tyler, 2007) and the Speech Learning Model (SLM; Flege, 1995, 2007) are the most cited for second language learning. The Second Language Linguistic Perception Model (L2LP; Colantoni et al., 2015; Escudero, 2005) will also be discussed briefly. The SLM, as well as the recently revised version (SLM-r; Flege & Bohn, 2021), is particularly relevant to the current study as it is more targeted to production, and it will therefore be discussed in detail below.

The Perceptual Assimilation Model (PAM; Best, 1994a, 1994b), was originally developed for explaining how non-native speech is perceived by naïve listeners, i.e., listeners who have no experience with the language they are perceiving. This model has been extended to L2 speech perception in the PAM(-L2) (Best & Tyler, 2007). The starting point is that when a listener is presented with an unfamiliar speech sound, this sound will likely be assimilated to the native phoneme that is most similar articulatorily to the unfamiliar sound. Depending on the structure of the native phonological system and characteristics of the unfamiliar speech sound, different forms of perceptual assimilation are predicted to occur, both regarding how the sound is identified and the degree to which the listener can discriminate phonological contrasts in the unfamiliar language. For L2 learning the focus is on contrasts, rather than on individual speech sounds as in the SLM which will be explained in more detail later in this section. In both models there is a common phonological space for L1 and L2 sounds and through perceptual learning existing sound categories can change and new categories can be formed. However, the models differ in some key assumptions about what is being perceived.

While the SLM assumes that listeners perceive acoustic cues in the speech signal, PAM assumes that the listener perceives articulatory gestures. In PAM it is the perceived degree of similarity between gestures perceived to be involved in the production of an L2 speech sound and gestures belonging to an L1 phonological category that decide perceptual assimilation, not acoustic phonetic similarity. For L2 listeners, these different types of assimilation can be used to predict whether the listener can perceive contrasts when they are first encountered, the likelihood that the listener will learn to distinguish L2 sounds, and the likelihood that new L2 categories will be formed. The different types of assimilation, and the predictions for perceiving contrasts will not be detailed here as production is the focus of the current study.

The L2LP model (Colantoni et al., 2015; Escudero, 2005), that has also been implemented computationally (van Leussen & Escudero, 2015), has some shared characteristics with both the SLM and PAM, but the L2LP is more explicit about the levels of processing and development stages involved in L2 (speech) learning. The model makes predictions from the onset of learning to higher levels of proficiency, unlike the SLM(-r) and the PAM-L2 which are more focused on the initial stages. Like PAM(-L2), the L2LP focuses on L2 perception and assimilation of contrasts, rather than individual speech sounds, but the L2LP assumes listeners perceive acoustic cues, like in the SLM(-r). A study using the L2LP framework has also shown that speakers of the same L1 might assimilate L2 sounds to different L1 categories, but they do so in a systematic way (Mayr & Escudero, 2010). This suggests that the different types of assimilation are not solely determined by the L1 but may also be affected by individual differences between speakers.

The SLM (Flege, 1995, 2007) describes L2 speech learning through the process of establishing (or failing to establish) new phonetic categories (i.e., category assimilation and category dissimilation). The SLM was developed to "account for age-related limits on the ability to produce L2 vowels and consonants in a native-like fashion" (Flege, 1995, p. 237). While it is often evoked to explain age related effects on L2 acquisition and related studies often focus on language learning after moving into an L2 environment, the model is also relevant for the current study because of its predictions of interactions between the phonetic subsystems of the L1 and L2, and the role of language input.

According to the SLM, unfamiliar speech sounds are processed as allophones of L1 categories in early stages of L2 learning. A new L2 phonetic category may eventually be established for the sound in the L2, depending on the perceived difference between the L1 and L2 sounds. For instance, English /θ/ and /ð/ may be perceived and produced as /t/ or /d/ by a Norwegian learner, but as the learner gains experience with the language a new L2 specific category may be formed and eventually produced (category dissimilation). The opposite might also happen, and the L2 sound continues to be processed as an instance of the L1 category (category assimilation). Importantly, even if two separate categories are not formed, exposure to perceptually similar (but not equal) L1 and L2 speech sounds may lead to the development of a "merged" category that is different from monolingual norms in both the L1 and L2.

Examples of different outcomes in speech learning can be found in Flege (1987), a study that investigated voice onset times (VOT) for /t/ and vowel formants for /u/ and /y/ produced by French-English bilinguals, three groups of English-French bilinguals, and French and English monolinguals. The English-French group with the least L2 experience produced English-like VOTs in both languages. The English-French group with intermediate L2 experience (more formal L2 education, but less exposure to spoken native French than the high experience group) produced more French-like VOTs than the least experienced group and the largest VOT difference between the languages overall. For both groups English VOTs did not differ significantly from English monolingual productions, and neither produced that were VOTs similar to the French monolinguals. The English-French group with the most L2 experience, and the similarly experienced French-English group, produced VOT values that were intermediate to the mean monolingual values in each language and significantly different from L1 monolingual norms, suggesting that a merged category had been developed. The value of the second formant (F2) in French /u/ produced by French monolinguals was significantly different from the mean F2 in all the bilingual groups, and none of the groups differed significantly from the monolingual production of English /u/. F2 difference measured between French /u/ and /y/ (/y/ present in French only) increased with experience for the English-French groups. The results suggest that the least experienced English-French group confused /u/ and /y/, but with more L2 experience the pronunciation starts to differentiate as there was a significant difference between F2 values in /u/ and /y/ for the more experienced English-French groups, even though these productions differed from the L2 norm.

In this view, L1 and L2 representations of speech sounds (including assimilated categories) are part of two separate phonetic subsystems, existing in a "common phonological space", which can mutually influence each other in perception and production. The influence of experience with the L2 could for instance be accounted for with exemplar theory, where representations are affected by the frequency and recency of perceived speech sounds. As mentioned, there is no consensus regarding the nature of representations and the link between perception and production in L1 processing. In Levelt's model of speech production (Levelt & Wheeldon, 1994), the units of perception were hypothesised to be abstract representations of overlearned syllables, not segments as in the SLM. Within that framework it is still possible to produce unfamiliar structures, but these would have to be computed. This framework can therefore be extended to accommodate an L2, where new structures may have to be computed initially and can become overlearned with time. One could imagine that L1 gestural scores are easily transferred or reused for L2 syllables if they are sufficiently similar. This could for instance be influenced by accent, the specificity of the abstract syllable, and the similarity of speech sounds. For instance, the syllable and quality of segments in the Norwegian-English cognate *bag* are quite similar (Norwegian: /bæg/ - English: /bæg/), but less so in the cognate compass (Norwegian: /kumˈpɑs/ - English: /ˈkʌmpəs/). At some point many speakers acquire and produce non-native speech sounds and clearly also higher-level characteristics such as stress and intonation. As such, a complete model of L2 acquisition could draw on several mechanisms proposed for L1 models.

Some aspects of the SLM have recently been updated with the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021). The primary aim has shifted from focusing on age-related limits on speech learning to accounting for "how phonetic systems reorganise over the life-span in response to the phonetic input received during naturalistic L2 learning" (Flege & Bohn, 2021, p. 23). With this change, the focus has also shifted from comparing groups of bilinguals that are more or less experienced in their L2, to look at the earlier stages and processes in L2 speech learning. It should be noted that the current study does not include a group of bilinguals that are directly referenced by the model, in addition to the fact that they are generally acquiring their L2 in an L1 environment.

The original SLM assumed perception preceded production in speech learning, and therefore that the accuracy of L2 production depends on the accuracy of L2 perception. There is evidence that shows a link between performance on tasks assessing perception and production accuracy (e.g., W. Baker & Trofimovich, 2006; Flege, 1999; Kim & Clayards, 2019; but see, Kartushina & Frauenfelder, 2014; Peperkamp & Bouchon, 2011), but these results only show that there is a link between the two, not directionality. Therefore, the SLM-r assumes that perception and production coevolve in speech learning. Evidence for a flexible perceptual system, and that there is a link between perception and production of non-native speech sounds, also comes from studies where participants train perception and/or production of non-native speech (e.g., Thorin et al., 2018; Zhang & Peng, 2017). Individual differences have also been reported. For instance, one study found that training perception improved production overall, however on the individual level there was perceptual learning with no production improvement as well as improvement in production but not perception (Bradlow et al., 1997). Kartushina and Frauenfelder (2014) did not find evidence of a link between perception and production performance. However, individual differences in variation within speech sound categories and their position in the F1-F2 vowel space were found to predict the accuracy with which L2 vowels were produced. Individual differences in L1 production, quantified by acoustic distances between vowel pairs, have also been found to affect the ability to discriminate L1 vowel contrasts (e.g., Franken et al., 2015; Fridland & Kendall, 2012).

The SLM-r hypothesises that the precision of L1 categories at the time of L2 speech learning onset will influence how easily the listener can perceive differences between L1 and L2 sounds. This in turn affects whether the learner can form a phonetic category for the L2 sound. Category precision is defined as "the variability of the acoustic dimensions measured in multiple productions of a phonetic category" (Flege & Bohn, 2021, p. 36). The authors go on to mention that this variation depends both on the degree of divergence between speech sounds in phonetic space and possibly individual differences in "auditory acuity, early-stage (precategorical) auditory processing, and auditory working memory" (Flege & Bohn, 2021, p. 36). The use of articulatory divergence measures between L1 and L2 sounds in the current study, as well as other measures in the phonology domain, may therefore shed light on the interaction between these variables in proficient bilinguals, as well as the effects on the word production process.

A final difference between the SLM and the SLM-r that may be relevant to current study is the role of L1 representations in the acquisition of L2 representations. The SLM postulated that the formation of L2 phonetic categories might be limited by features and weighting of acoustic cues present for speech sounds in a learner's L1. As a result, a phonetic category formed for an L2 sound might be represented differently than monolingual native representations of the same sound. In the SLM-r, on the other hand, they adopt the full access hypothesis (see e.g., Escudero & Boersma, 2004) to account for findings such as L2 speakers being able to access and acquire features not used in their L1 (e.g., Iverson & Evans, 2007). However, this is also subject to individual differences. A study comparing the use of spectral and durational cues to the Dutch low vowel contrast /aː/ - /ɑ/ by L1 Dutch, L1 Spanish and L1 German speakers found that both native language and language experience affected perception and the reliance on either acoustic dimension (Escudero et al., 2009). The L1 Spanish listeners had experience with Dutch as an L2, and even though their L1 only has one low vowel and no durational contrasts, their performance on a vowel categorisation task was similar to that of the Dutch native speakers. This was not the case for the L1 German speakers with no Dutch language experience, whose vowel inventory is more similar to Dutch. Overall, the L1 Spanish speakers relied more on durational cues than the L1 Dutch speakers, however a minority of the L1 Spanish group relied more on spectral information. These results show that learners can acquire speech sounds and cues that are not part of their L1 and highlight the importance of considering individual differences.

To summarise, in the SLM(-r) L1 and L2 representations of speech sounds exist in a common phonological space and are part of two separate phonetic subsystems which can mutually influence each other. L2 speech learning is thought to be influenced by L1, but the effects on perception, learning, and formation of new L2 categories are subject to individual differences. This includes the prediction that individuals with more precise L1 categories at the onset of L2 speech learning will be more successful at perceiving differences between L1 and L2 sounds, and forming phonetic categories for L2 sounds.

6.2.3 Sound structure and L2 processing: Speech, cognates, and switching

Up to this point this chapter has concentrated on describing how speech sounds are perceived, what the units of perception might be, and models of perception and production, all of which assume non-switched language production. However, as detailed in Chapter 5, bilinguals often switch between their languages, and evidence suggest that there is non-selective language activation during speech production as shown by both switching and cognate effects. This section will focus language production tasks and the acoustic output. Two studies looking at the acoustic output in language switching tasks have been identified (Goldrick et al., 2014; Olson, 2013). These will be described, before discussing a study investigating cognate effects in articulation using a different methodology.

Goldrick et al. (2014) investigated voice onset times (VOTs) produced by Spanish-English bilinguals in a language switching task. Specifically, VOTs in the /t/-/d/ contrast, which is realised differently in Spanish and English. All participants also spoke Catalan where the VOT contrast is similar to that in Spanish. The study additionally manipulated cognate status and whether initial phonemes on sequential trials were the same or different. Both voiced and voiceless sounds in the nondominant language English were affected in the direction of a smaller VOT difference (decreasing the contrast between the languages) on switch trials compared to stay trials. In voiceless stops they also found a stronger effect for cognates compared to noncognates. These findings are explained with interactive activation in word production (see e.g., Section 2.3). That is, cognates in the nontarget language are more activated than noncognates, and through cascaded activation phonetic processes receive more activation as well, leading to a stronger influence on the acoustic output.

Olson (2013) also studied VOTs in a switching task with bilingual speakers of Spanish and English, one English-dominant group and one Spanish-dominant group. Three different trial structures were used to create two monolingual language contexts (one with 95% L1 trials and 5% L2 trials, and one with 95% L2 trials and 5% L1 trials) and a bilingual context with half of the trials in each language. In the monolingual condition switch trial VOTs were significantly different from stay trials and the difference was asymmetric across languages. While L1 productions were significantly affected in the direction of L2 norms on switch trials for both groups, L2 VOTs were similar across trial types. This is opposite to the finding in Goldrick, Runnqvist and Costa (2014), however the trial structure was different. In the bilingual condition, where the trial structure resembles other language switching studies, VOTs were not affected by switch status in either language. Both of these studies found an effect of more accented speech on switch trials compared to stay trials. However, it is unclear how this interacts with language dominance, as these two studies report the three possible outcomes: more accented speech in the L1, more accented speech in the L2, and no difference between the two.

Cognate effects on the acoustic realisation of speech have also been observed in studies using other methodologies. Amengual (2012) found a cognate status effect on Spanish VOTs produced in a monolingual Spanish context by reading target words embedded in a carrier sentence. Four groups of bilinguals speaking Spanish and English (Heritage Spanish, Heritage English, English L1 – Spanish L2 and Spanish L1 – English L2) all produced significantly more English-like VOTs (longer) for Spanish cognate words than Spanish noncognate words, while no significant difference was found for the Spanish-Catalan bilingual control group. Interestingly, VOTs collapsed over condition fell within the monolingual range for all groups and the VOT differences between the groups were not significant. This could suggest that the different groups with different bilingual profiles, have established L2 phonetic categories, in the sense described by the Speech Learning Model described above, that are similar to the ones produced by simultaneous bilinguals.

These results suggest that cognate effects are robust across several groups of bilinguals. However, the study also found an influence of individual differences. In three of the four groups (not the Spanish heritage speakers), there were some individuals who did not produce a significant difference between cognates and noncognates. In the Spanish-Catalan group none of the participants produced a significant difference between cognates and noncognates. Amengual (2012) proposes an extension of existing exemplar models to account for cognate effects in bilingual articulation. As described above, in exemplar theory for speech perception, heard instances of words and sounds are stored as exemplars and organised in relation to each other based on similarity. For bilinguals, then, instances of cognates, in both of their languages, would be stored close to each other, if not overlapping. In production, exemplars in an activated region contribute to the articulation plan, and this could account for cognate effects on articulation, and why noncognates are not affected in the same way.

The results described above suggest that the acoustic characteristics of speech output are affected by cognate status and therefore that cross-language activation can influence the whole speech production process, including articulation. As shown in this chapter there are different theoretical views and models for explaining representation of sound structure and processing in both L1 and L2. The most relevant to the current study is the Speech Learning Model(-r), in which L2 speech learning is described as a process of establishing (or failing to establish) new phonetic categories. L1 speech sounds and those established for the L2 are thought to exist in a common phonological space where they can influence each other. In the revised version of the model, learning outcomes are hypothesised to be determined by the precision of L1 categories at the onset of L2 learning.

This thesis explores the relationship between language production and acoustic output from a different angle, using individual difference measures of articulatory divergence between the articulations of similar L1 and L2 sounds. The SLM predicts that individuals with more precise L1 phonetic categories are more likely to perceive a difference between similar L1 and L2 speech sounds, and in turn more likely to establish representations for L2 speech sounds both in perception and production. The claim is not that there is a direct relationship between perception and production, but that these domains are closely linked. In production, then, a larger acoustic difference between similar L1 and L2 speech sounds, would suggest more precise categories than for those individuals who exhibit smaller differences. These measures are added as predictors to the previously reposted models of picture naming and language switching. As several studies report effects of individual differences, the divergence measures will also be compared to the language tests and factors extracted from the bilingual profile questionnaire. In the next section methodological considerations and the methods for obtaining the divergence measures will be discussed.

6.3 Method and methodological considerations

This section starts with a discussion of methodological considerations and methods for measuring vowel formants and articulatory divergence. Then English and Norwegian vowel inventories are compared, before the vowels of interest in the current study are described. The next section details the development of a forced aligner. In the final section, formant measures from the aligned vowels are evaluated and refined. Formant values were processed and visualised in R (R Core Team, 2022) and figures were made using the packages ggplot2 (version 3.4.2; Wickham, 2016) and cowplot (version 1.1.1; Wilke, 2020).

6.3.1 Measuring articulation and divergence

Spoken L2 proficiency often focuses on the degree of accentedness or comprehensibility and there are many ways of assessing proficiency in this domain, as discussed in Section 4.2.3. Typical methods include comparing L2 productions to a native norm or using native speaker ratings of non-native productions. In the current study, participants are Norwegian L1 speakers who predominantly acquired their L2 English in Norway. As such, L2 learning has occurred in an environment where they have been exposed to both non-native English, as well as varieties of native English. The participants reported having different target accents in their L2[23] and are therefore not easily compared to native speakers of any variety of English, nor L2 speakers of English immersed in an English-speaking country. Quantifying proficiency by obtaining ratings from native English speakers or comparing their phonetic articulations to a native English norm is therefore challenging in the current study. Instead, the current study employs a measure of articulatory divergence. That is, the acoustic difference between individual participant's productions of similar, but not identical, L1 and L2 speech sounds.

To measure speech divergence in the current study, several speech sounds that could be challenging for this group of speakers and suitable for measuring individual differences were considered. For Norwegian L1 speakers, several English segments may be challenging, such as the dental fortis fricative and lenis fricatives in general. The dental fortis fricative is a challenging new consonant for

[23] Number of participants and self-reported accent: 23 American or mostly American, 9 British or mostly British, 11 a mix of two or more accents (British, American, Australian, Scottish, Norwegian), 9 none in particular/depends on interlocutor, 5 Norwegian/Scandinavian, and 3 other.

Norwegian speakers learning English, however many acquire target-like pronunciation after a while. There are no voiced fricatives in Norwegian (Nilsen, 2010), and therefore voicing can remain challenging for advanced speakers. Assessing voicing in fricative and affricate minimal pairs in general, such as /s/ - /z/, could therefore be an option for the relatively proficient speakers in the current study. However, this would be less suitable for a continuous individual differences measure as voicing and duration interact in signalling these contrasts and it is unclear which acoustic measures would be most appropriate for this group of speakers. Instead, vowel quality was measured in L2 vowels that remain challenging for proficient speakers and the most similar L1 vowels. The difference between two sets of similar, but not identical, L1 and L2 vowels serve as continuous measures for investigating individual differences (detailed in Section 6.3.2 below). Therefore, this section will focus on measuring vowels, and options for measuring degree of overlap or divergence in vowels.

Vowel quality is typically assessed acoustically by measuring formants, or resonance frequencies, measured in Hertz (Hz). Vowel quality is mostly determined by the first formant (F1) and the second formant (F2). These values determine the placement of vowels within a two-dimensional vowel space. The F1 is mostly affected by the height of the tongue body, so that that vowels produced with a low tongue body have a high F1 and those produced with a high tongue body have a low F1. The F2 is mostly affected by horizontal placement of the highest point of the tongue, where front vowels have a higher value and back vowels have a lower value. Other aspects that contribute to the perception of vowels and may vary in production include the third formant (F3), formant dynamics and temporal aspects, but these will not be the focus of this section.

There are several options for calculating the difference between the formant values to determine the degree of divergence. As mentioned above, the first two formants (F1 and F2) values are central to perceived vowel quality. In combination they determine the placement of vowels in the vowel space, therefore only measures which take both F1 and F2 into consideration will be considered. In this thesis it is of particular interest to measure the degree of (dis-)similarity between L1 and L2 vowel segments produced by individual speakers, rather than comparing these L2 segments to a native norm. This section will therefore also include measures typically used in sociolinguistic research. These measures are for instance used for measuring ongoing vowel mergers and splits, and changes in phonetic production after moving to a new area.

In L2 research, the Mahalanobis distance score or the quadratic distance has been used to assess how L2 productions differ from native targets (Kartushina et al., 2016; Kartushina & Frauenfelder, 2014) and comparing groups of bilinguals (Yang & Fox, 2017). This measure was not used in the current thesis as it has mainly been used to compare L2 productions with native norms, however it has been used in one recent study of individual differences in the production of an L2 contrast (Melnik-Leroy et al., 2022). Euclidian distance has been used as a measure of distance between vowels in studies on both L1 (e.g., Kendall & Fridland, 2012) and L2 production (e.g., Amengual, 2016b; Bion et al., 2006). In a review of options for measuring dialectal mergers, two of the evaluated measures are Euclidian distance and adjusted Euclidian distance (Nycz & Hall-Lew, 2014). These were evaluated in terms of (amongst other criteria) their ability to capture distance and overlap between speech sound categories and how they cope with unbalanced data. The benefits of using the Euclidian distance measure include it being simple to calculate and transparent in the sense that the distance can be reported in Hertz. A potential drawback is that it only provides information about the distance between values, not the degree of spread or overlap between speech sound categories. Using the adjusted Euclidian distance measure and mixed effect regression makes it possible to control for unbalanced data, but it does not provide an overlap measure.

A measure that has been frequently used in sociolinguistic research is the Pillai score, first used to investigate vowel overlap in New Zealand English (Hay et al., 2006). The authors argue that this is better than using Euclidian distance, as the Pillai score reflects the overlap between distributions in F1-F2 space. It also allows for capturing variation from F1 and F2 in one measure, and it is robust against violations of normality and homogeneity (Nycz & Hall-Lew, 2014). The Pillai-Bartlett statistic, or Pillai score, represents the "proportion of one variance that can be predicted by another variance, given any known conditioning" and is an output of a Multivariate Analysis of Variance (MANOVA) (Hall-Lew, 2010, p. 2). The Pillai score provides a value between 0 and 1, where 0 signifies the most overlap between two vowels and 1 signifies the largest distance or divergence. The Pillai score was also found to provide the best accuracy and precision in a study reviewing four vowel overlap measures using simulated data (Kelley & Tucker, 2020).

While several studies have used the Pillai score and found that aspects are preferable to other measures quantifying differences between vowels, Johnson (2015) argues that the Pillai score is not a direct measure of distance or overlap. This critique, and others, were also brought up by Nycz and Hall-Lew (2014), mentioned above, who argue that it captures distance and overlap only in terms of "overall difference" and that it is not a clear measure of neither distance nor overlap. Johnson (2015) also reports that the Pillai score is less reliable for unequal sample sizes and certain distributions of data, but that most these of issues can be avoided if one uses Bhattacharyya's Affinity (BA) instead. BA ranges from 1, signifying complete overlap, to 0, signifying no overlap. The BA is additionally more sensitive and will reach 0 faster when the distributions are clearly divergent compared to the Pillai score (Johnson, 2015).

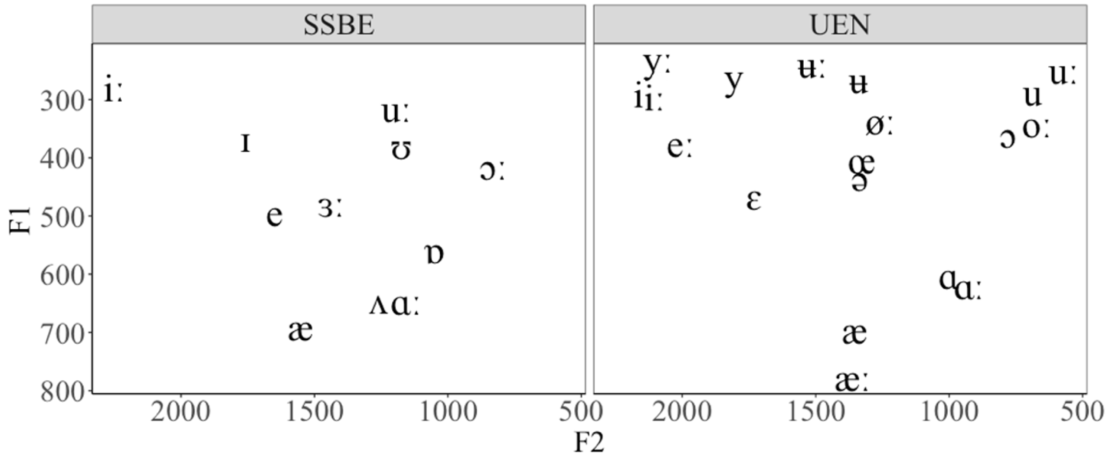
In the current project, both the Pillai score and BA seem appropriate for capturing the divergence between pairs of L1 and L2 speech sounds that may overlap to a varying degree in individuals' productions. The literature suggests that there are benefits to using both of these measures and they should, in theory, provide similar results. While the Pillai score has been used more frequently, the BA seems to be more reliable. One sociolinguistic study reports finding similar results using the Pillai score and BA (Labov et al., 2016). It is unclear how these measures will affect the analysis. Therefore, both will be compared to bilingual profile measures and language test scores and entered as predictors in models of the speech production data analysed in previous chapters. While the measures employed here measure the difference between L1 and L2 productions of similar, but not identical, vowels, it should be kept in mind that this measure does not address how similar or different L2 vowel productions are to vowels produced by native English speakers. The next section will provide an overview of Norwegian and English vowel inventories, and details about the segments of interest in the current study.

6.3.2 Comparing Norwegian and English vowels

When selecting stimuli for the picture naming task, part of the selection criteria was to include English words containing segments that can be challenging for Norwegian speakers, and Norwegian words containing segments that were similar to these challenging L2 segments (/dʒ tʃ v w z s θ ɜː uː ʊ ʌ/). In general, this contributed to making a balanced and challenging set of pictures for the participants to name. For the divergence measure I selected a subset of stimuli and chose two L1-L2 vowel contrasts. Both languages have relatively complex vowel systems as can be seen in the F1-F2 vowel spaces plotted in Figure 21. Two sets of vowels that are similar, but not identical, across the two languages were selected as these were expected to be challenging and suitable for assessing individual differences, even for relatively proficient speakers. As will be detailed below, the selection of vowels for the divergence measure was not just based on acoustic similarity between L1 and L2 vowels, but also known errors or substitutions that L1 Norwegian speakers tend to make in their production of L2 English.

The first L2 vowel of interest is the long close back monophthong /uː/. Norwegians often articulate this sound with a quality similar to the Norwegian vowel /ʉː/, which is more fronted than the English target sound. Another possible substitution is Norwegian /uː/ which has a closer and more back quality than the English target sound. The quality of the English sound is therefore said to lie somewhere "between" the two Norwegian sounds (Nilsen, 2010). In addition, both Norwegian sounds are produced with more lip rounding than the English counterpart. Nilsen and Rugesæter (2015) argue that the most frequently used replacement is /ʉː/ and attribute this to the relationship between Norwegian spelling and pronunciation. That is, words spelled with <u> in Norwegian tend to be pronounced with an /ʉː/ sound, while words with the letter <o> are pronounced with an /uː/ sound. As there are two potential L1 influences on the articulation of English /uː/, formant measures from both L1 sounds were compared to L2 productions before deciding on the first L1-L2 contrast for the divergence measure. I will refer to these as /uː/-type vowels.

Figure 21
English and Norwegian F1-F2 vowel spaces



Note. English values obtained from 5 male speakers of Standard Southern British English (SSBE) as reported by Deterding (1997). Norwegian values were obtained from one speaker producing vowels with an Urban East Norwegian (UEN) accent (Kristoffersen, 2000). F1 and F2 measured in Hertz.

The second L2 vowel of interest is the English /ʌ/, a short open central monophthong which tends to be "very troublesome" for native Norwegian speakers (Nilsen, 2010, p. 112). There are two L1 vowels, /œ/[24] and /ɔ/, Norwegians often use instead of the target L2 vowel /ʌ/, even though the quality of the Norwegian vowel /ɑ/ would be more similar to English /ʌ/. This is also attributed to spelling as the Norwegian vowel /ɑ/ is always found in words written with <a>, while the spelling of English words pronounced with an /ʌ/ vary. In English words containing <u> or <ou>, Norwegians tend to produce vowels similarly to native /œ/, and for words written with an <o> Norwegians tend to use Norwegian /ɔ/. In addition to quality differences in F1 and F2, English /ʌ/ is produced without lip rounding, while /œ/ and /ɔ/ are rounded vowels (Nilsen, 2010; Nilsen & Rugesæter, 2015). As for English /u:/, there appears to be two potential L1 speech sounds that may influence the articulation of the target L2 sound, therefore both were analysed before deciding on the second L1-L2 contrast. These will be referred to as /ʌ/-type vowels.

The paragraphs above describe possible L1 substitutions for the two selected L2 speech sounds. These L2 sounds may of course also be produced with a quality

[24] Nilsen (2010) uses the symbol /ø/. Following Kristoffersen (2000), I use /œ/ for the short vowel and /ø:/ for the long vowel in Norwegian.

that is more similar to native English speakers' articulations of /uː/ and /ʌ/, and therefore more different from L1 articulations. When comparing L1 and L2 vowels, speakers' whose productions show larger difference between L1 on L2 sounds, i.e., more divergent, may therefore produce more target-like segments and have more precise phonetic categories (in the SLM-r sense, as discussed in Section 6.2.2), than speakers' whose L1 and L2 productions are less divergent. As shown above, L2 articulation may be influenced by L1 speech sounds. For /uː/-type vowels, based on the descriptions above, the most variation might be observed in F2. For English /ʌ/, F1 and F2 may vary both in magnitude and direction of difference depending on the L1 influence. Compared to English /ʌ/, the L1 /œ/ has a more high and front articulation, and productions of /ɔ/ might be more back and possibly a little higher. Using measures such as the Pillai score and BA therefore seem suitable, both considering individual differences in category precision, and to capture variation in F1 and F2. The next section will detail the development of a forced aligner for Norwegian and English speech.

6.3.3 Forced aligner for segmenting bilingual speech

In the current study, the speech materials are vowels taken from words produced in the picture naming and language switching tasks. In order to segment the speech data, a forced aligner was developed to allow for automatic segmentation of the words and specifically the vowels of interest to the current project[25].

There are several available tools for English and other languages, such as the web-based aligner DARLA (Reddy & Stanford, 2015). However, these were not suitable for the current project as there is no control over the dictionary and language models. As there was no existing aligner for Norwegian, we trained an automatic forced aligner on L1 Norwegian and L2 English speech for this project, using tools implemented in other aligners. In the current project, the same speakers are producing both L2 English and L1 Norwegian speech and the majority of the models detailed below were trained on both languages simultaneously. Since the goal in the current project is to compare vowel quality across languages, using the same aligner and same language model avoids bias from using different aligners or different training data.

[25] Developed in collaboration with Jan Zandhuis in the Experimental Linguistics Lab at the University of Agder.

The input data consisted of all speech from correct trials on the picture naming and language switching tasks (described in Chapter 5). The single words produced during these tasks should be reasonably well aligned with phonemes and have clear word boundaries due to silences on either side of the word. The recordings were converted to mono and a 16000 Hz sample rate. The speech was produced by 60 native speakers of Norwegian with English as their strongest L2 (44 female, 16 male) and a mean age of 24.10 years (*SD* = 4.25). A more detailed description can be found in Section 4.4.1, where the participants in question are referred to as the experimental group. As detailed below, the first few models were run with data from 4 participants. We included all words produced by the speakers, rather than just the ones containing the vowels of interest, to provide a richer training set. A subset was then used for the acoustic analysis of articulatory divergence (detailed in Section 6.3.4). We trained our model with the Montreal Forced Aligner (MFA; McAuliffe et al., 2017) that uses Kaldi for speech recognition (Povey et al., 2011). Dictionaries were created for the Norwegian and English words in the input data. The annotation of phonemes was based on ARPABET, but with some additions to distinguish sounds in the two languages where necessary. There were six main iterations of the model, the first five with MFA version 1.0.1, while the final version was trained with version 2.0.0a9.

With each iteration, Praat (Boersma & Weenik, 2021) TextGrids with the alignments produced by the model were compared to manual measurements for a subset of the data. The first three versions were run on both Norwegian and English speech data from four speakers (7154 words in total). Evaluation of the model and adjustments focused on improving the detection and alignment to word onsets and offsets. Initially, the training algorithm was quite sensitive to noise, clicks and creaks. Thirty alignments were checked manually, all of which had misaligned onsets leading to skewed alignment of words, and in the worst cases no alignment to actual speech. This was both due to word onsets being aligned to noise and failures to align low intensity initial segments. The results improved when editing the input dictionary to include a noise silence at both ends of the phoneme description. In addition, we shortened the sound files leaving about 50 ms of silence before and after each word. This improved the alignment of word onsets and offsets for all participants (118 files checked manually), however unvoiced plosives were not always aligned correctly for two of the four speakers. From here on out, the models were trained on data from all speakers.

The fourth model was trained on Norwegian and English speech simultaneously. In general, this produced better word onset alignments than the three first versions. 216 words (108 from each language) were segmented manually and the average difference between manual and automatic onset alignment was 9 ms. Plosive onset detection was improved, but the model struggled with initial fricatives. The alignment of the vowels of interest for divergence analysis was checked in 84 samples from one male and one female speaker. Specifically, the analysis focused on whether the vowel was properly aligned, and which part of the segment contained the stable portion of the vowel. For 42 vowels the stable portion was in the first half of the segmented interval, 27 had a stable portion in the middle, 4 in last half, and the rest had no stable portion. In addition, there was missing data where the model had failed to produce TextGrids.

The participants in the current experiment were included on the basis of being L1 Norwegian speakers with English as their strongest L2. While they all were native speakers of Norwegian, they came from different dialect backgrounds. Stimuli were transcribed phonologically according to Urban East Norwegian (UEN) production. Words that were produced with a vowel quality, syllable number or stress placement that differed from UEN were coded during the experiments. In the last rounds of model training, these productions were excluded from the training set (774 words excluded). While all productions were included initially to maximise the training set, it is possible that the model would be more successful when some variation was removed. With this new dataset we first trained two models separated by language, one for Norwegian and one for English. There were still some missing output files, but this time only from the switching task. At this point 231 words were checked manually focusing on the alignment of the vowels of interest. Overall, the alignments were good, but the alignments of vowels were skewed in certain contexts, specifically those articulated preceding or following nasal or liquid segments. However, the stable portion of the vowel varied in a predictable manner. We then ran the model again on both Norwegian and English data combined. This time different TextGrids were missing. Further inspection revealed that the missing files were dropped during training. After installing a new version of the aligner (MFA v2.0.0a9) this was no longer an issue.

The final version, using the new aligner, was trained on both languages simultaneously. We stored the language model after training but discarded the output. We then ran the speech files again with the dedicated aligner and the stored language model to obtain the best and final TextGrids. From the final model, 463 forced alignments (3.74%) of words containing the vowels of interest were checked manually. This included the 231 words checked for the previous version of the model, words with triphone boundary shifts exceeding 30 ms from the previous model, and additional words containing the vowels of interest. The new aligner performed the best at picking up word onsets and ignoring noise. As with the previous version, the stable portion in the vowels of interest varied in a predictable manner. Therefore, the final model was able to successfully align segments overall, allowing for automatic segmentation of the relevant speech data, but some adjustments were necessary for measuring formant values which will be detailed below.

6.3.4 Formant analysis

The evaluation and development of the forced aligner focused on whether the vowels of interest were reliably aligned to the vowel segment in the speech data, and whether a stable portion of the vowels could be identified for measuring formants. There are several options when choosing a location for formant measurements, including measuring formants at the intensity peak (Bergmann et al., 2016) or the temporal midpoint (Mairano et al., 2019; Melnik-Leroy et al., 2022), using measures averaged over a predefined portion of the vowel (Barreda, 2021), measuring at the highest F2 value (Hay et al., 2006), or using different measurement points depending on the of type vowel (Labov et al., 2013). A study comparing different locations of measurements of formants in read speech produced at both a normal and fast speaking rate found that the compared methods essentially produced the same output (van Son & Pols, 1990, p. 1692). It should be noted that there is very likely more variation in the data presented in this thesis compared to the study above. The analysed speech was produced by one speaker who was a newscaster, in contrast to data from two languages and a group of speakers in the current study. However, it is still of interest that results were comparable regardless of whether formants were measured at the midpoint of the vowel, averaged for the whole vowel, or based on heuristics identifying stable portions of the vowel.

In the current study, speech samples were taken from the picture naming and switching experiments. Six words (3 cognates and 3 noncognates) were selected for each vowel of interest, and each word was produced 6 times by each speaker (12960 tokens in total). Words that had been excluded in the RT analysis in Chapter 5, were also removed from this analysis. In addition, 30 Norwegian tokens were removed where dialectal differences lead to the participant not producing the target vowel. For example, the Norwegian word *kurv* /kʉrʋ/ (English *basket*) was sometimes produced with the vowel /ɔ/. This left 12232 tokens in total for the formant analysis. The stimulus materials in the picture naming and switching experiments were selected to meet a number of conditions. Therefore, the environments surrounding the vowels of interest are not controlled. The evaluation of the aligner focused on identifying stable portions of the vowel (in the segmentation provided by the model) to ameliorate potential effects of coarticulation.

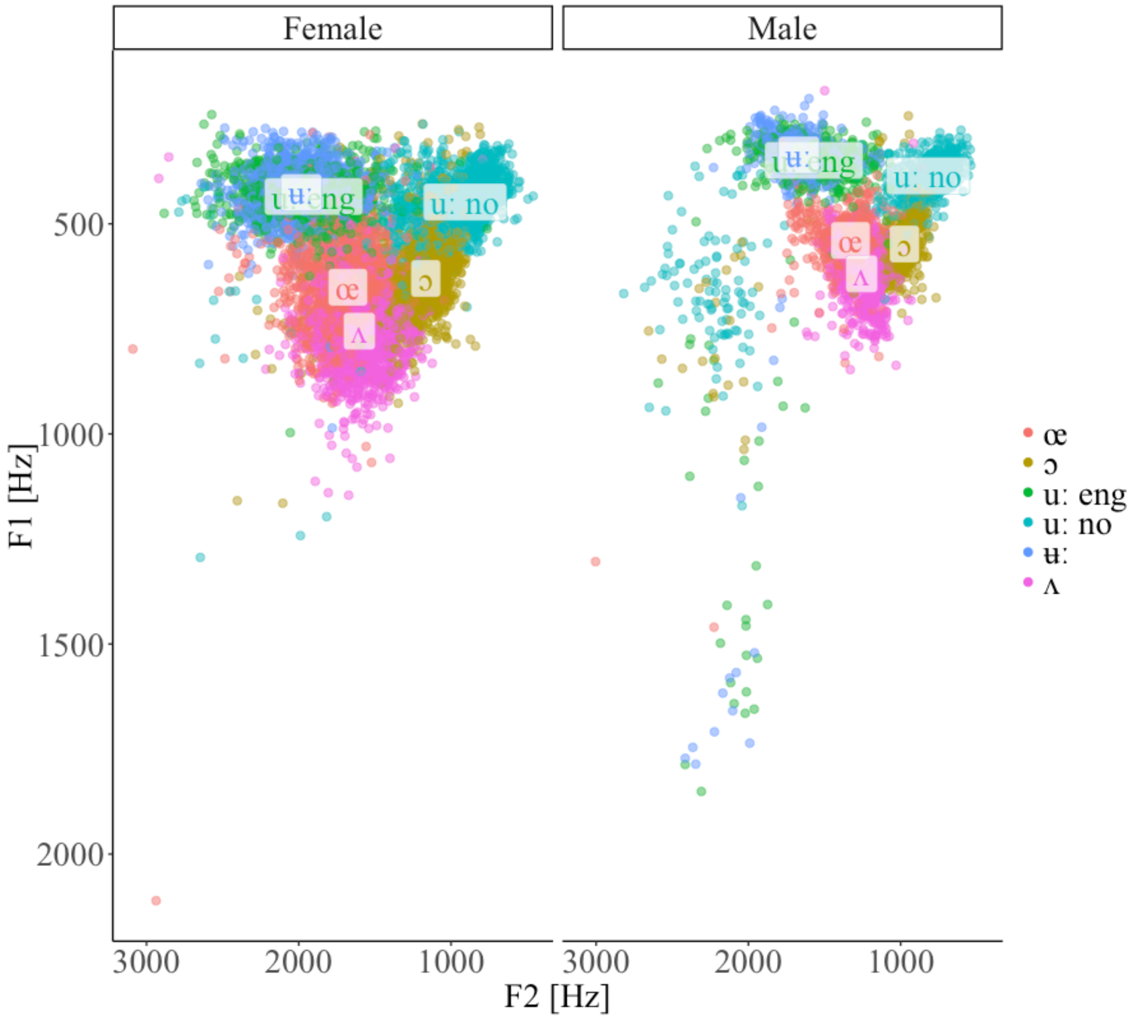
The current study uses formant frequencies measured in Hertz. Since numeric frequency differences between formants are not necessarily perceived as different by the human auditory system, many studies use a psychoacoustic scale instead (e.g., Bark units or the mel scale) that more accurately reflects auditory perception. Values were not transformed in the current study as the interest lies in differences in production and not perception. The recordings were not normalised prior to the formant analysis, even though this is common in phonetic analysis, as each person is their own control and normalisation might mask individual differences in production (Barreda, 2021).

Formants were measured in Praat (Boersma & Weenik, 2021) using a modified version of the TB-Track Vowels script by Brato (2016). Recordings were downsampled to 11025 Hz and then the Linear Predictive Coding (LPC) formant tracks were logged using a 0.01 time step, 0.025 window length, and pre-emphasis from 50 Hz. The maximum formant for male speakers was set at 5000 Hz and at 5500 Hz for female speakers. Formant measures were taken at 11 equidistant points in the aligned vowel from 0% to 100% of the vowel duration. For vowels with a stable portion in the middle, the final measures of F1 and F2 was the median of measurements taken at five points from 30% to 70% percent of the vowel duration. For vowels with a stable portion in the first half of the vowel measurements were taken from 10% to 50% of the vowel duration (see Appendix M for an overview).

As mentioned above, two sets of three vowels were selected for the analysis, /uː/-type vowels (L2 /uː/, L1 /ʉː/, and L1 /uː/) and /ʌ/-type vowels (L2 /ʌ/, L1 /ɔ/, and L1 /œ/). The formant measures were first inspected visually, and the data were cleaned before deciding on the vowel pairs for the divergence measures. Figure 22 shows two vowel spaces (female voices on the left) with all raw data points. Vowel type is indicated by the colour of the dots. While the overall pattern shows the expected distribution of vowel sounds (with individual differences) there is a clear tail for the male speakers. In an effort to reduce the tail, observations exceeding 2.5 standard deviations from the by-talker by-vowel mean were removed, following Melnik-Leroy et al. (2022). The remaining 11288 data points are plotted in Figure 23, which shows that the trim improved vowel spaces overall. The upper limit was reduced for both F1 and F2, and the distribution of vowel points was less dispersed compared to the raw data in Figure 22. Though the tail observed in the male data was reduced after the trim, it was still an issue. To investigate this issue further vowel plots divided by speaker were produced. The individual plots for female speakers can be found in Appendix N, and the plots for the male speakers are shown in Figure 24.

Figure 22

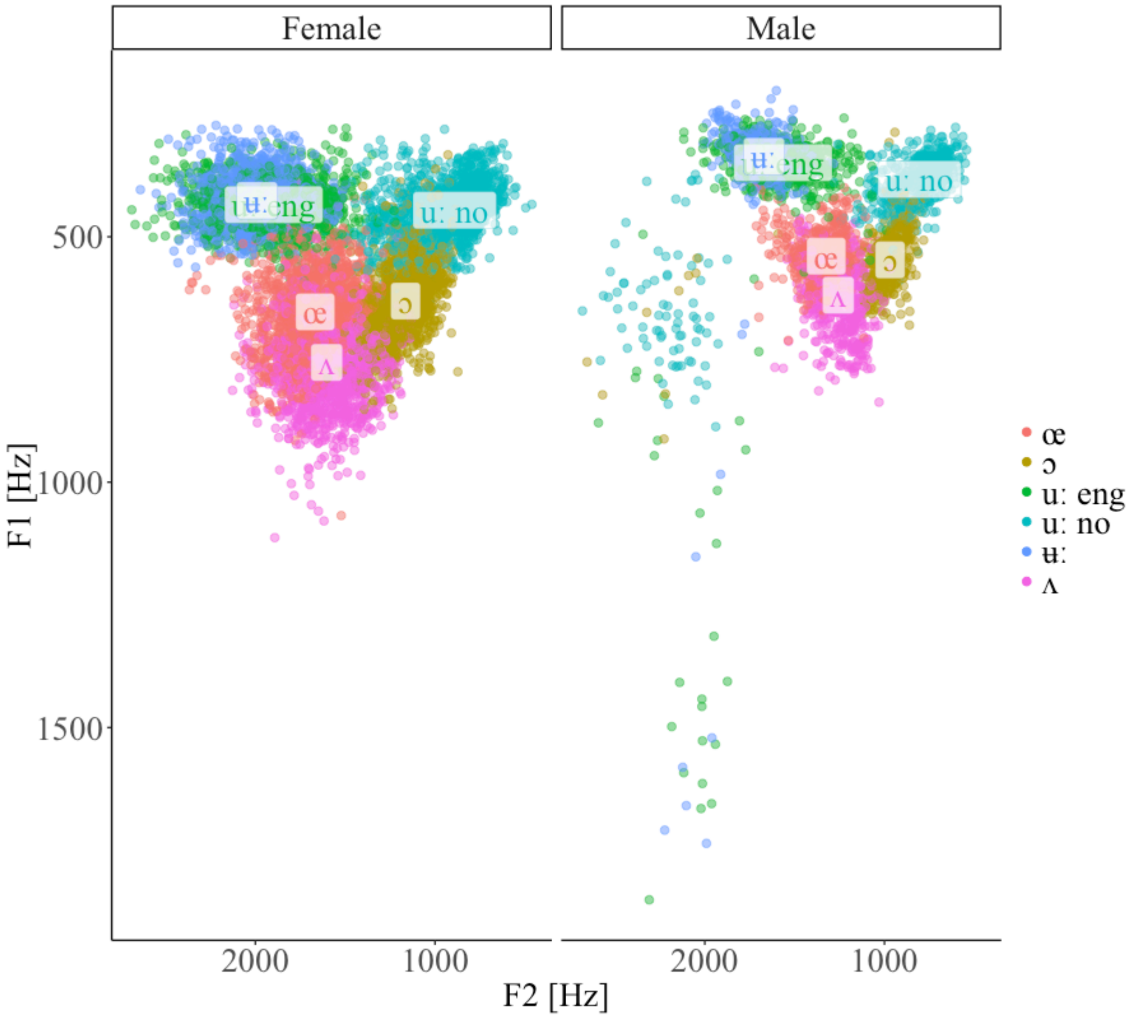
F1-F2 vowel spaces with raw data - individual data points and vowel means



Note. Vowel type indicated by colour. "u: eng" refers to English /u:/ and "u: no" refers to Norwegian /u:/.

Figure 23

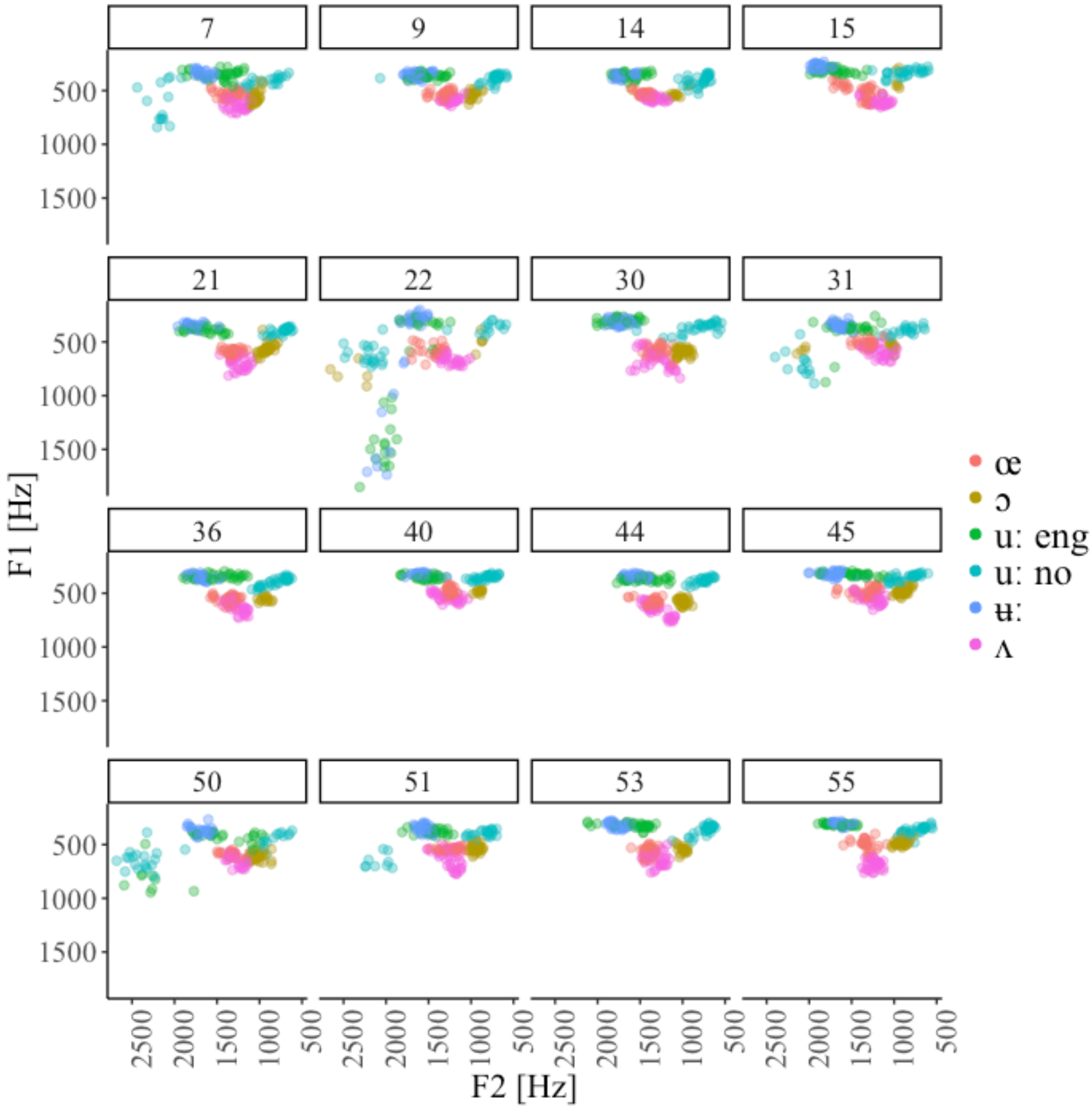
F1-F2 vowel spaces after +/- 2.5 SD trim - individual data points and vowel means



Note. Vowel type indicated by colour. “u: eng” refers to English /u:/ and “u: no” refers to Norwegian /u:/.

Figure 24

Individual F1-F2 vowel spaces for male speakers after +/- 2.5 SD trim

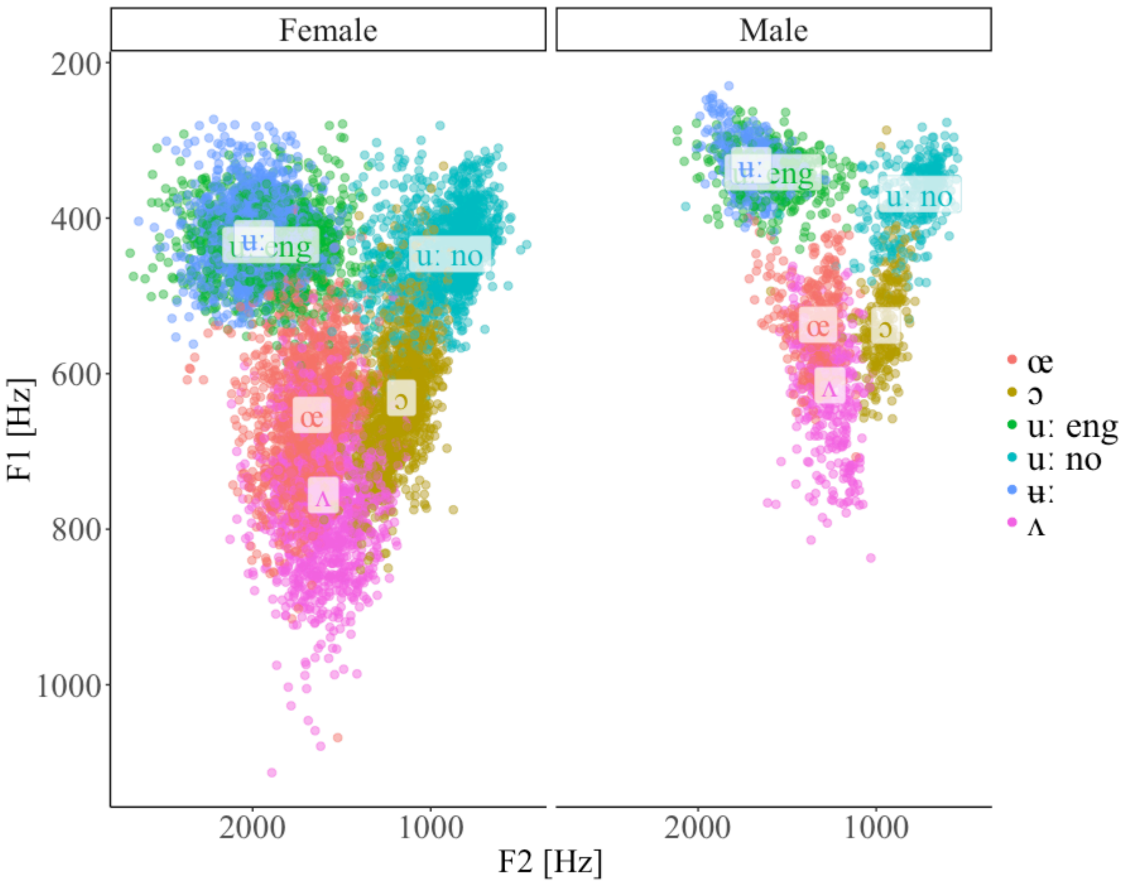


Note. Vowel type indicated by colour. "u: eng" refers to English /u:/ and "u: no" refers to Norwegian /u:/.

The individual speaker plots and participant means revealed that the problems stemmed from /u:/-type segments. While one male speaker accounted for most of the tail, measurements were inconsistent for 5 male speakers in total (7, 22, 31, 50 and 51 in Figure 24). These observations were not linked to specific words containing /u:/-type segments. The next step was to inspect the sound files from which these measurements were taken. Two main issues were identified. First, the presence of creaky voice in some of the vowel segments led to the first formant not being recognised and the value of the second formant was erroneously assigned to the first formant. Additionally, some productions exhibited unusual pitch movement on the vowel which seemed to interfere with the formant analysis. These observations were all measurement errors. Other studies with manual segmentation or manual correction of automatic segmentation report correcting errors manually (e.g., Strange et al., 2001). Since all measurements were done automatically, both the segmentation and the extraction of the formant values, it was not justified to manually correct the formant values from the five participants mentioned above. No clear option was identified for fixing these issues with the /uː/-type segments, and it was not desirable to keep participants with incomplete data. Therefore, the participants in question were dropped from the analysis. The final data is visualised in Figure 25. The rest of the analyses were completed with 55 native speakers of Norwegian with English as their strongest L2 (44 female, 11 male). The mean age was 23.87 years (*SD* = 4.26), and they had 16.27 years of education on average (*SD* = 2.33).

Figure 25

F1-F2 vowel spaces for the final dataset with individual data points and vowel means

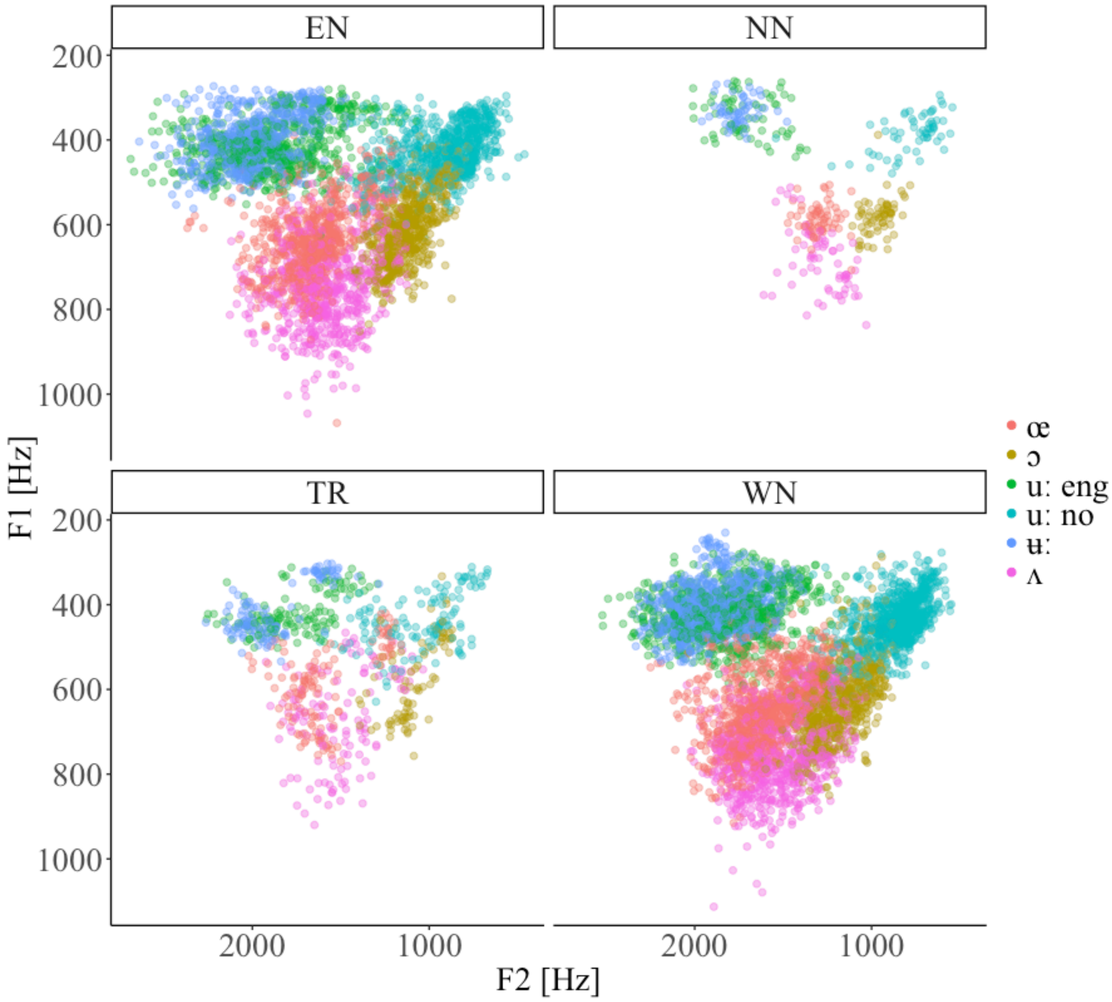


Note. Vowel type indicated by colour. "uː eng" refers to English /uː/ and "uː no" refers to Norwegian /uː/.

The data were checked to see if there were any large differences between L1 dialect groups. Productions where the target vowel was not produced were excluded prior to analysis (see Section 6.3.3). To assess whether any large differences were visible between dialects, the observations were plotted by dialect group. The participants self-reported their Norwegian dialect and the specificity of their responses varied. Using four large dialect groups, following Papazian and Helleland (2005, pp. 84–85), the distribution was as follows: 22 speakers of Eastern Norwegian (*Østnorsk*), 27 speakers of Western Norwegian (*Vestnorsk*), 4 speakers from the middle of Norway (*Trøndersk*) and 2 speakers of Northern Norwegian (*Nordnorsk*). The largest regional group contained 23 speakers of Southern dialects (included in the Western group above). Figure 26 shows separate vowel plots for each dialect group (plots separated by speaker gender and dialect group are available in Appendix O). The groups are uneven in size, but overall, the distributions within the F1-F2 vowel space were relatively similar and the data were not trimmed any further.

Figure 26

F1-F2 vowel spaces for the final dataset - divided by L1 dialect groups



Note. Eastern Norwegian (EN), Northern Norwegian (NN), from the middle of Norway (TR) and Western Norway (WN). Vowel type indicated by colour. "u: eng" refers to English /u:/ and "u: no" refers to Norwegian /u:/.

6.4 Results

This section starts by describing the final articulatory divergence measures. These measures are then compared to self-ratings of pronunciation proficiency and language test scores, before they are compared to bilingual profile factors (all from Chapter 4). Finally, the divergence measures are added as predictors in models of latencies from the picture naming and language switching task. All data processing and statistical analysis of the data was conducted in R (R Core Team, 2022). The alpha was set at 0.05, but for model reduction in Section 6.4.4., the alpha was set at 0.1 for retaining model terms.

6.4.1 Divergence measures

Based on the final dataset and the visualisation of the vowels of interest, two Norwegian speech sounds were selected for the divergence measures. While the symbol /uː/ is typically used for phonemic transcriptions of the vowels found in English *tooth* and Norwegian *stol*, the visualisation shows that these speakers clearly do not use Norwegian /uː/ for English /uː/. The more phonetically similar sounds are L1 /ʉː/ (e.g., in *hus*) and L2 /uː/. Therefore, the final contrast for the /uː/-type divergence measure was L2 /uː/ and L1 /ʉː/. For the /ʌ/-type vowels (L2 /ʌ/, L1 /ɔ/ and L1 /œ/) there is more overlap between English /ʌ/ and Norwegian /œ/, making the final contrast for the /ʌ/-type vowels L2 /ʌ/ and L1 /œ/.

As previously discussed, two different measures for quantifying divergence were selected for this study, the Pillai score and Bhattacharyya's Affinity (BA). These measures were calculated in R, following Stanley (2019). BA was calculated using the adehabitatHR package (version 0.4.19; Calenge, 2006). Input formant measures (F1 and F2) for each vowel token were the median of 5 measures taken from the stable portion of the vowel (details in Section 6.3.4). After removing 5 participants the total number of observations was 9975, including 6880 observations for vowels in the final contrasts. For each speaker there was a maximum of 36 tokens per vowel due to the design of the study (M = 31.27, SD = 2.31). In total, four divergence measures were computed per participant, both a Pillai score and BA for the each of the vowel pairs /uː/-/ʉː/ and /ʌ/-/œ/. With the Pillai score values closer to 0 indicate overlap while values closer to 1 indicate divergence, while BA ranges from 1 indicating overlap to 0 indicating divergence. Overall, productions of the /uː/-type vowels were more overlapping (Pillai: M = 0.24, SD = 0.18; BA: M = 0.73, SD = 0.14) than the /ʌ/-type vowels (Pillai: M =0.47, SD = 0.18; BA: M = 0.64, SD = 0.14).

6.4.2 Divergence measures, self-rated pronunciation proficiency, and language test scores

The association between self-ratings of L1 and L2 accent and each divergence measure was assessed with separate correlations for each language and each divergence measure. The scale provided for rating proficiency ranged from 0 (*none*) to 10 (*perfect*) (see Table 4 in Section 4.4.1 for full scale specification). Self-ratings of proficiency pronouncing L2 (M = 7.20, SD =1.45) met the assumptions for correlation, but self-ratings of proficiency pronouncing L1 (M =

9.58, *SD* = 0.79) failed to meet the assumptions of linearity and normality. Unsurprisingly there was a larger range of ratings in L2 (3-10) than L1 (6-10). For the /uː/-/ʉː/ contrast, linearity and normality was not met for the Pillai score. Normality was met for both the Pillai score and BA for the rest of the contrasts, but linearity was an issue for the /ʌ/-/œ/ BA. As for the language test comparisons in Section 4.4.3, several transformations were attempted without sufficiently improving the data. Therefore, the analysis was run with non-transformed data and both Pearson and Kendall correlation measures are reported. There were no significant results (see Appendix P, Table P1).

Multiple linear regressions were run with the divergence measures and L1 and L2 performance on four language tests (spelling, vocabulary, elision, and serial nonword recognition). These are described in detail in Sections 4.3.2 – 4.3.5. Norwegian and English test scores correlated highly and therefore the linear models for each of the divergence measures against the test scores were separated by language. All values were centred prior to analysis. The assumptions for the residuals were met for most of the linear models. However, the residuals for the English test scores and the /ʌ/-/œ/ BA divergence measure were not normally distributed (W = 0.96, p = .044). Re-running the regression with a log transformation improved normality (W = 0.97, p = .225), but this did not change the results. All results are printed in Appendix P, Table P2.

A higher score on the L1 serial nonword recognition task was significantly related to more divergent productions of the /uː/-/ʉː/ contrast as measured by both the Pillai score (Estimate = 0.64, *SE* = 0.25, t = 2.60, p = .012) and BA (Estimate = -0.60, *SE* = 0.18, t = -3.36, p = .002). There was a significant positive relationship between more divergent /ʌ/-/œ/ productions, measured by the Pillai score, and L2 spelling (Estimate = 0.26, *SE* = 0.12, t = 2.07, p = .043). Finally, there was a marginal effect in the same direction for the /uː/-/ʉː/ Pillai scores and L2 spelling (Estimate = 0.24, *SE* = 0.13, t = 1.89, p = .065). There were no other significant results.

6.4.3 Divergence measures and bilingual profile factors

The relationship between the divergence measures and the 10 factors extracted from the bilingual profile questionnaire (detailed in Section 4.4.2) was addressed with four multiple linear regressions. The bilingual profile factors do not correlate highly, but Pillai scores and BA do, so separate multiple linear regressions were run for each divergence measure with the 10 factors. All values were centred prior to the analysis and the assumptions for the residuals were met.

For the /uː/-/ʉː/ pair, the regression equations were not significant for the Pillai score ($F(10,44) = 1.395$, $p = .214$), nor the BA ($F(10,44) = 1.369$, $p = .226$). Adjusted R^2 was .068 and .064, respectively. There were no significant effects for either divergence measure. For the /ʌ/-/œ/ pair, there were significant effects for both divergence measures (printed in Table 14). Significant regression equations were found for both the Pillai score ($F(10,44) = 3.081$, $p = .005$) and the BA ($F(10,44) = 3.382$, $p = .002$). Adjusted R^2 was .278 and .306, respectively.

Table 14
Results for factors significantly related to divergence measures

	Pillai score /ʌ/-/œ/				Bhattacharyya's Affinity /ʌ/-/œ/				VIF
Factor name	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	
L2 interaction and mixing	0.053	0.024	2.25	.029	-0.233	0.018	-1.31	.197	1.12
Language development	-0.078	0.029	-2.69	.010	0.054	0.022	2.47	.018	1.34
L2 exposure and mixing	0.078	0.025	3.11	.003	-0.063	0.019	-3.36	.002	1.35
L1 informal exposure	0.085	0.035	2.45	.019	-0.071	0.026	-2.73	.009	1.89
Informal learning	-0.033	0.026	-1.32	.194	0.041	0.019	2.11	.041	1.19

Note. Non-significant results in grey.

More divergent productions were significantly related to higher values on the factor *L2 interaction and mixing* for the Pillai score, but not the BA. For the BA, but not the Pillai score, less divergent productions were related to higher values on the *Informal learning* factor. For the three other factors there were similar significant effects for both divergence measures. More divergent productions were associated with reaching language development milestones at an earlier age, as captured by the *Language development* factor. Higher values of both the *L2 exposure and mixing* and *L1 informal exposure* factors were also associated with more divergent productions.

6.4.4 Divergence measures, picture naming, and language switching tasks

Finally, the divergence measures were centred and entered as predictors in linear mixed models of latencies in the picture naming and language switching task to assess whether articulatory divergence modulates the speed of speech of processing. There was no evident speed accuracy trade off in the Chapter 5 analysis, therefore accuracy will not be addressed here. The input data and maximal models were identical to the analyses in Chapter 5, apart from the predictors. The models were fitted using the lme4 package (version 1.1.28; Bates et al., 2015). For the models of the switching data, the random effects structure was reduced until the model converged. This was not necessary for the picture naming models. The model terms were evaluated using the drop1 function. Naming latencies from the picture naming and switching task were modelled separately for each divergence measure, giving a total of 8 models. Benjamini-Hochberg significance (Benjamini & Hochberg, 1995) with a critical value of 0.05 is reported to control the false discovery rate for multiple comparisons. The results are only reported for the three models where the divergence predictor was retained after model reduction. Residual plots for these three models and a table with Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC) for each step in the model reduction can be found in Appendix Q.

The /ʌ/-/œ/ Pillai score and /uː/-/ʉː/ BA predictors were not retained in the models of reaction times from the picture naming task. There were no significant main effects of the /uː/-/ʉː/ Pillai score or the /ʌ/-/œ/ BA, but they both appeared in significant interactions. The /uː/-/ʉː/ Pillai score interacted significantly with naming language (results in Table 15). In both languages faster naming was associated with more divergent productions, but the effect was larger in L1 (trend: -.0.185, *SE* = 0.0917) than L2 (trend: -0.060, *SE* = 0.0923).

The BA /ʌ/-/œ/ predictor was involved in a significant three-way interaction with cognate status and naming language (Table 16). English and Norwegian reaction times for both cognate and noncognate words appear to be differently affected by the divergence measure. For English, the difference between cognate reaction times and noncognate reaction times are smaller when the vowels are more divergent, while the opposite pattern emerges for the Norwegian words. There is a negative trend for noncognates named in L1 (-0.017, *SE* = 0.129), suggesting that L1 noncognates are named more slowly with more divergent productions, while more divergent productions are associated with faster naming for L1 cognates (trend: 0.039, *SE* = 0.144), L2 noncognates (trend: 0.061, *SE* = 0.131) and L2 cognates (trend: 0.013, *SE* =0.128). Submodels by language[26] show a significant effect of cognate status in both languages (English: Estimate = 0.094, *SE* = 0.027, t = 3.54, p <.001; Norwegian: Estimate = 0.101, *SE* = 0.025, t = 3.99, p <.001), but the interaction between cognate status and the BA score is marginal for English (Estimate = 0.064, *SE* = 0.035, t = 1.85, p = .070) and not significant for Norwegian (Estimate = -0.077, *SE* = 0.045, t = -1.70, p = .095).

In the language switching task the only divergence measure retained in the model was the /uː/-/ʉː/ Pillai score (Table 17). There was one significant two-way interaction between the /uː/-/ʉː/ Pillai divergence measure and naming language, however this was no longer significant after performing the Benjamini-Hochberg procedure.

[26] -1000 * 1/RT ~ CognateStatus*cbhatt.ʌ.œ +
(CognateStatus |SUBJ_ID) + (1|Item)

Table 15
Model output from the analysis of picture naming RTs with the Pillai /uː/-/ʉː/ divergence measure as a predictor

Model: -1000 *1/RT ~ CognateStatus + NamingLanguage + Session + cpillai.uː.ʉː + CognateStatus:Session + NamingLanguage:Session + NamingLanguage:cpillai.uː.ʉː + (NamingLanguage + CognateStatus + CognateStatus:NamingLanguage|SUBJ_ID) + (NamingLanguage + Session + NamingLanguage:Session|Item)

	Model summary			Model comparison			B-H
	Estimate	*SE*	*t*	χ^2	df	*p*	sign*
Fixed effects							
(Intercept)	-1.3018	0.0210	-61.93				
CogStatus	**0.1061**	**0.0235**	**4.51**	**26.46**	**1**	**<.001**	yes
Language	**-0.0292**	**0.0127**	**-2.30**	**5.12**	**1**	**.024**	yes
Session	0.0047	0.0107	0.44	0.0002	1	.989	no
cPillai /uː/-/ʉː/	-0.1226	0.0879	-1.40	0.64	1	.423	no
CogStatus:Session	**0.0390**	**0.0097**	**4.00**	**14.73**	**1**	**<.001**	yes
Language:Session	**0.2043**	**0.0692**	**2.95**	**8.00**	**1**	**.005**	yes
Language:cPillai /uː/-/ʉː/	**-0.1253**	**0.0549**	**-2.29**	**4.69**	**1**	**.030**	yes
Random effects	s^2						
Item (Intercept)	0.014						
Subject (Intercept)	0.017						

*Benjamini–Hochberg significance

Note. CogStatus = Cognate status (Cognate vs. Noncognate); Language = Naming language (English vs. Norwegian); Session (Session 1 vs. Session 2); cPillai /uː/-/ʉː/ (centred Pillai scores for the /uː/-/ʉː/ vowels). Significant results in bold.

Table 16

Model output from the analysis of picture naming RTs with the BA /ʌ/-/œ/ divergence measure as a predictor

Model: -1000 *1/RT ~ CognateStatus + NamingLanguage + Session + cBA.ʌ.œ + CognateStatus:NamingLanguage:cBA.ʌ.œ + CognateStatus:NamingLanguage + CognateStatus:Session + NamingLanguage:Session + CognateStatus:cBA.ʌ.œ + NamingLanguage:cBA.ʌ.œ + (NamingLanguage+CognateStatus+ CognateStatus:NamingLanguage|SUBJ_ID) + (NamingLanguage+Session+ NamingLanguage:Session|Item)

	Model summary			Model comparison			B-H
	Estimate	*SE*	*t*	χ^2	df	*p*	sign*
Fixed effects							
(Intercept)	-1.3023	0.0212	-61.31				
CogStatus	**0.1056**	**0.0236**	**4.48**	**26.50**	**1**	**<.001**	yes
Language	**-0.0293**	**0.0131**	**-2.24**	**5.12**	**1**	**.024**	yes
Session	0.0032	0.0112	0.28	0.004	1	.952	no
cBA /ʌ/-/œ/	0.0241	0.1266	0.19	0.83	1	.362	no
CogStatus:Language	0.0036	0.0165	0.22	0.04	1	.834	no
CogStatus:Session	**0.0373**	**0.0099**	**3.88**	**14.69**	**1**	**<.001**	yes
Language:Session	**0.2019**	**0.0699**	**2.90**	**6.95**	**1**	**.008**	yes
CogStatus:cBA /ʌ/-/œ/	-0.0039	0.0276	-0.14	0.03	1	.857	no
Language:cBA /ʌ/-/œ/	-0.0260	0.0756	-0.34	0.26	1	.607	no
CogStatus:Language:cBA /ʌ/-/œ/	**-0.1036**	**0.0419**	**-2.47**	**5.78**	**1**	**.016**	yes
Random effects	s^2						
Item (Intercept)	0.014						
Subject (Intercept)	0.018						

*Benjamini–Hochberg significance

Note. CogStatus = Cognate status (Cognate vs. Noncognate); Language = Naming language (English vs. Norwegian); Session (Session 1 vs. Session 2); cBA /ʌ/-/œ/ (centred Bhattacharyya's Affinity for the /ʌ/-/œ/ vowels). Significant results in bold.

Table 17

Model output from the analysis of language switching RTs with the Pillai /uː/-/ʉː/ divergence measure as a predictor

Model: -1000 *1/RT ~ CognateStatus + TrialType + NamingLanguage + Session + cpillai.uːeng.ʉː + CognateStatus:TrialType:NamingLanguage + CognateStatus:TrialType: Session + CognateStatus:TrialType + CognateStatus:NamingLanguage + TrialType:NamingLanguage + CognateStatus:TrialType + TrialType:Session + CognateStatus:TrialType:cpillai.uːeng.ʉː + CognateStatus:cpillai.uːeng.ʉː + TrialType:cpillai.uːeng.ʉː + NamingLanguage:cpillai.uːeng.ʉː + (CognateStatus + NamingLanguage + TrialType|SUBJ_ID) + (1|Item)

	Model summary			Model comparison			B-H
	Estimate	*SE*	*t*	χ^2	df	*p*	sign*
Fixed effects							
(Intercept)	-0.9773	0.0246	-39 .79				
CogStatus	**-0.0473**	**0.0223**	**-2.11**	**9.55**	**1**	**.002**	yes
TrialType	**0.0804**	**0.0077**	**10.47**	**88.25**	**1**	**<.001**	yes
Language	**0.0430**	**0.0096**	**4.46**	**8.57**	**1**	**.003**	yes
Session	-0.0311	0.0095	-3.27	4.55	1	.033	no
cPillai /uː/-/ʉː/	-0.0569	0.1082	-0.53	0.37	1	.543	no
CogStatus:TrialType	-0.0247	0.0103	-2.41	2.99	1	.084	no
CogStatus:Language	**-0.0326**	**0.0102**	**-3.21**	**8.16**	**1**	**.004**	yes
TrialType:Language	-0.0191	0.0104	-1.83	0.81	1	.368	no
Cognate1:session1	0.0236	0.0102	2.32	1.41	1	.235	no
TrialType:session1	0.0236	0.0104	2.27	1.17	1	.279	no
Cognate1:cPillai /uː/-/ʉː/	-0.0178	0.0364	-0.49	0.31	1	.575	no
TrialType:cPillai /uː/-/ʉː/	0.0046	0.0316	0.14	2.96	1	.085	no
Language1:cPillai /uː/-/ʉː/	-0.0842	0.0409	-2.06	4.08	1	.043	no
CogStatus:TrialType:Language	0.0243	0.0145	1.68	2.83	1	.093	no
CogStatus:TrialType:Session	-0.0305	0.0145	-2.11	4.45	1	.035	no
CogStatus:TrialType:cPillai /uː/-/ʉː/	0.0715	0.0405	1.77	3.12	1	.078	no
Random effects	s^2						
Item (Intercept)	0.005						
Subject (Intercept)	0.020						

*Benjamini–Hochberg significance

Note. CogStatus = Cognate status (Cognate vs. Noncognate); Trial type (Stay vs. Switch); Language = Naming language (English vs. Norwegian); Session (Session 1 vs. Session 2); cPillai /uː/-/ʉː/ (centred Pillai scores for the /uː/-/ʉː/ vowels). Significant results in bold.

6.5 Discussion

In this chapter two measures of articulatory divergence, the Pillai score and Bhattacharyya's Affinity (BA), were calculated for two pairs of similar L1 and L2 vowels. The divergence measures were compared to self-reported and tested proficiency, bilingual profile factors, and latencies from the picture naming and language switching tasks.

The first research question posed at the beginning of this chapter asked: Does L1-L2 speech divergence relate to self-rated pronunciation proficiency? This was addressed by correlating each divergence measure to self-ratings of pronunciation proficiency in L1 and L2. There were no significant results and no evidence of a relationship between self-rated pronunciation proficiency and the divergence measures. The quality of segments is, however, only one of several factors that contribute to the accentedness of speech along with stress and intonation (see e.g., Section 4.2.3 of this thesis). Additionally, the extent to which L1 Norwegian speakers are aware of the differences between the L1 and L2 vowels considered here is unclear, and therefore the divergence between L1 and L2 vowels may not be something they consider in their self-ratings of accent.

After the experiments, participants (N = 60) were asked if there were any L2 English speech sounds they found difficult to pronounce (65.0% responded yes) and if they had noticed any English speech sounds that were difficult to pronounce for other Norwegians (73.3% responded yes). If so, they were asked to indicate which speech sounds were difficult. One person reported specifically finding L2 /uː/ difficult to pronounce, and the rest of the responses included one or more challenging consonants or consonant clusters[27], specific words (N = 3), or did not provide an answer (N = 2). There were generally more responses regarding which L2 speech sounds other Norwegians may find difficult, and these responses also mainly concerned consonants[28]. Three respondents mentioned English intonation as being the most difficult for Norwegians when speaking English. There were two responses each for "vowels" in general, L2 /uː/ and L2 /ə/. In addition, there were single mentions of /ɒ/, /ɛ/, and three diphthongs. Considering these responses, the

[27] Difficult L2 consonants, contrasts, and clusters participants report finding difficult to pronounce, in descending order: 9 /ɹ/, 6 /w - v/, 5 /z - s/, 4 /θɹ/ or /ɹθ/, 3 /tʃ/, 3 /θ/, 2 /dʒ/, 2 /kw/, 1 /ʃ/, 1 sequential liquids

[28] Difficult L2 consonants noticed for other Norwegian speakers, in descending order: 26 /θ/, 17 /w - v/, 13 /ɹ/, 9 /z - s/, 7 /θɹ/ or /ɹθ/, 4 /ð/, 4 pronouncing silent letters, 3 /dʒ/, 3 sequential liquids, 2 /ʃ - ʒ/, 1 /tʃ/

lack of correlations between the articulatory divergence measures and self-reported proficiency are not surprising, as participants tended to focus on other aspects of the speech signal rather than vowels.

The next two research questions asked: Does L1-L2 speech divergence relate to objective language test scores? and How do L1-L2 speech divergence measures relate to aspects of bilingual profile? The first question was addressed by running multiple regressions with each divergence measure and the test scores, separated by language. Participants completed a vocabulary task, a spelling task and two tasks assessing phonological processing, elision and serial nonword recognition (SNWR). Both the Pillai score and BA for the /uː/-type vowels were significantly related to L1 SNWR scores. The Pillai score for the /ʌ/-type vowels was significantly related with the L2 spelling scores and the /uː/-type Pillai score was marginally related to L2 spelling scores. The direction of the effects suggested that more divergent productions were associated with better performance on the tests in question. There were no significant effects for vocabulary or elision scores in either language, L2 SNWR scores and L1 spelling scores.

In Chapter 4 we saw that self-ratings of pronunciation correlated with elision scores for both Norwegian and English, but not with SNWR scores. One explanation put forward for the discrepant results of the two phonological tests was that the two tests measure different aspects of phonological memory. The elision task, which involves manipulating and producing speech segments, might be more related to the participants' current language behaviour, while the SNWR measures aspects of phonological processing and memory that may be more important in early language development. L2 SNWR scores were found to correlate with the *Language development* factor, suggesting that participants who reported reaching L1 and L2 language milestones at an earlier age performed better on the L2, but not the L1, SNWR task. In the current chapter, more divergent L1 and L2 productions of the /uː/-type vowels, measured by both the Pillai score and BA, were related to higher L1 SNWR scores. There were no significant relationships between the divergence measures and L2 SNWR scores or elision scores in either language.

The L1 and L2 /uː/-type vowels, as produced by participants in the current study (Section 6.4.1) and native speakers of Norwegian and English (Section 6.3.2), are closer in F1-F2 space than the /ʌ/-type vowels. The revised Speech Learning Model (SLM-r), discussed in Section 6.2.2, hypothesises that the precision of L1 categories at the time of L2 speech learning onset will influence how easily the listener can perceive differences between similar L1 and L2 sounds, and in turn learn to produce the L2 sounds. In this view, better performance on the L1 SNWR task may be related to more precise L1 categories, which in turn increases the likelihood that participants produce the L2 /uː/-type sound with a different quality than the L1 /uː/-type sound. Considering these results and those obtained in Chapter 4, phonological memory measured by the SNWR task in these young adult proficient speakers, seems to be related to specific phonological abilities that are associated with early language learning.

None of the phonological tests were significantly related to divergence measures for the /ʌ/-type vowels. As described in Section 6.3.2, Norwegian speakers tend to substitute English /ʌ/ with vowel qualities similar to Norwegian /œ/ or /ɔ/, probably due to influences from spelling, even though the more acoustically similar sound would be Norwegian /ɑ/. The /ʌ/-type divergence measures were calculated with L1 /œ/ and L2 /ʌ/. These sounds are less similar acoustically than the /uː/-type vowels and therefore productions may be less influenced by individual differences in category precision. L2 spelling scores significantly correlated with the /ʌ/-type Pillai score and marginally correlated with the /uː/-type Pillai score. The divergence measures were to some extent related to tested proficiency in the domains of phonology and spelling. The clearest effect was found for /uː/-type divergence and the L1 SNWR score, where both the Pillai score and the BA showed a similar effect. The observed effects are interesting and consistent with close relationships between phoneme and grapheme representations.

The relationship between the divergence measures and bilingual profile was addressed by running multiple regressions for each divergence measure against the 10 factors extracted from the bilingual profile questionnaire (Section 4.4.2). Unlike the regressions with the test scores, there were no significant effects of either divergence measure for the /uː/-type vowels. However, there were significant results for both divergence measures of the /ʌ/-type vowels. Both the Pillai and BA divergence measures were associated with *Language development*, *L2 exposure and mixing*, and *L1 informal exposure*. In addition, the Pillai score was related to *L2 interaction and mixing*, and the BA was related to *Informal learning*. There were no significant effects involving the five other factors.

The relationship between the *Language development* factor and the divergence measures suggests that the participants who reported reaching language development milestones earlier have more divergent productions of the /ʌ/-type vowels. The /ʌ/-type Pillai measure also correlated with L2 spelling, and the analysis in Chapter 4 found that the participants who reached language milestones at an earlier age performed better on the L2 SNWR task and the L1 elision task. Taken together this could suggest that what lies behind more divergent productions of the /ʌ/-type vowels is a more language general aptitude in the phonological domain, compared to the effects observed for the /uː/-type vowels and L1 SNWR above. However, with this interpretation it is not clear why similar effects were not found for the *Language development* factor and the elision and spelling tasks in both languages.

The variables in the *L2 exposure and mixing* factor, that was significant for both the Pillai score and BA, are similar to those in the *L2 interaction and mixing* factor, that was significant for the Pillai score only. Both entail more exposure to L2 in general, but the former contains less L2 exposure through communication and the latter is associated with less accented English and more L2 exposure from friends and family. Higher scores on both of these factors are associated with more divergent productions of the /ʌ/-type vowels. This is in accordance with the SLM(-r) which is explicit about the importance of (communicative) language exposure, in addition to the precision of L1 categories, for L2 speech learning. However, these factors were not related to /uː/-type divergence. It is possible that for the more acoustically similar /uː/-type vowels, divergent production is more dependent on fine-grained individual differences in L1 category precision or auditory acuity compared to the /ʌ/-type measures.

Norwegian speakers tend to use vowels similar to L1 /œ/ or L1 /ɔ/, in place of L2 /ʌ/, rather than the more acoustically similar L1 /ɑ/, and this is likely influenced by spelling, as discussed previously. Participants with more exposure to L2 input, as measured by the factors, might rely more on auditory input and be less influenced by spelling, thereby producing L2 /ʌ/ with a quality that is more different from L1 /œ/. In the current study the /ʌ/-type divergence measures are based on the difference between L1 /œ/ and L2 /ʌ/, and the interpretations above remain speculative. A future, more targeted study might more systematically address the influence of individual differences in language exposure vs. category precision by including more vowels, for instance comparing productions of L2 /ʌ/ to L1 /ɔ/, /ɑ/, and /œ/. The hypothesis that L1 SNWR scores and individual differences in L1 category precision are more likely to affect acoustically similar sounds, might for instance be examined further comparing productions of L2 /ʌ/ and L1 /ɑ/.

The *L1 informal exposure* factor was related to both /ʌ/-type divergence measures and suggests that more informal exposure to the native language in daily life (through television, music and reading) and more L1 learning through music, television and self-instruction is associated with more divergent representations. This factor also has a negative loading of L2 exposure from family. Participants indicated the relative amount of time being exposed to L1 and L2 from various sources, therefore more exposure in L1 generally means less exposure in L2. This seems to conflict with the effects of the two L2 factors described above, which suggested that participants with more exposure to L2 input produced more divergent /ʌ/-type segments. However, there are also aspects of L1 and L2 exposure that are not captured by this factor, such as language mixing and language exposure from friends which may provide L2 exposure.

The final significant effect was an association between the *Informal learning* factor and the /ʌ/-type BA measure. Higher values on the *Informal learning* factor were associated with less divergent production. This factor includes variables relating to informal learning for both languages, in terms of a higher contribution of music and tv for L2 learning, and L1 learning (in order of magnitude) from reading, music, television, school and friends. In addition, there is a negative association to L1 learning from family and a positive loading on L2 accent being perceived as non-native. While the L2 accent question has one of the lowest loading values in the factor, the association between more accented speech and less divergent representations falls in line with predictions. As mentioned, the SLM-r is explicit about the importance of communicative exposure for L2 speech learning. The highest positive loadings in this factor relate to L2 language learning through non-communicative exposure. This could suggest that these participants were less exposed to L2 through communicative input, and as such this association between the factor and the divergence measure is in line with the predictions of the SLM-r. However, variables associated with L1 learning also load positively onto this factor so this interpretation only partially accounts for the relationship between the factor and divergence.

As discussed in Chapter 4, the five first factors found in the factor analysis were generally the most reliable. Only two of these, *L2 exposure and mixing* and *Language development*, were related to divergence measures. The *L2 accent and interest* factor contains positive loadings of variables such as finding L2 accent important, interest and effort in improving L2 accent, ability to imitate accents and paying attention to pronunciation. However, this did not correlate with any of the divergence measures. Speech segments participants in the current study recognise as difficult to pronounce, both for themselves and others, was described when discussing the lack of a relationship between self-ratings of accent and the divergence measures. This showed that participants mainly focused on consonants, and may therefore not be aware of, or interested in, the production of these specific vowels. It is possible that a significant relationship could have been observed if a different aspect of speech production had been measures, but in the current study no such relationship was found. The divergence measures also did not relate to the *General L2 proficiency* and *General L1 proficiency* measures, which supports the observation that pronunciation and accent tends to be separated from other linguistic abilities (e.g., Jilka, 2009).

To summarise, the L1-L2 speech divergence measures relate to aspects of bilingual profile to some extent, but only for the /ʌ/-type divergence measures. More divergent productions, measured both with the Pillai score and BA, relate to reaching language development milestones at an earlier age and more L2 exposure and mixing. More divergent productions measured by the Pillai score are related to more L2 interaction and mixing. Finally, more informal learning, which seems to entail less communicative L2 exposure, is associated with less divergent productions as measured by the BA. These findings support a role for individual differences in phonological abilities or aptitude and L2 exposure for L2 speech learning. However, there is also an association between informal exposure to the L1 and more divergent productions which is not readily interpretable.

Considering the relationship between the divergence measures, factors, and test scores, it is clear that the two divergence measures (the Pillai score and BA) are not always comparable. The results also differ depending on which L1-L2 vowel pair is used for calculating the divergence measure, i.e., the /uː/-type vowels or the /ʌ/-type vowels. The results seem to indicate that individual differences in the production of the more acoustically similar /uː/-type vowels are more related to specific phonological abilities, while individual differences in the production of /ʌ/-type vowels are more related to variables associated with L2 exposure. However, there is only partial support for these interpretations of the difference between the /uː/-type and /ʌ/-type vowels.

The final research question in this chapter asked whether the degree of articulatory divergence was related to language production behaviours in the picture naming and language switching tasks. The divergence measures were added as predictors into models of the picture naming and switching task latencies. In the picture naming task models, the /uː/-type Pillai score and the /ʌ/-type BA appeared in significant interactions, but the /ʌ/-type Pillai score and /uː/-type BA predictors were not retained in the models. There was a two-way interaction between naming language and the /uː/-type Pillai score. In both languages, less divergence is associated with slower responses, while more divergence is associated with faster responses. However, the positive effect of more divergent production was stronger for L1.

The /ʌ/-type BA interacted with cognate status and naming language. In English the difference between cognate and noncognate latencies was smaller with more divergent vowels, while the effect was opposite for Norwegian words. L1 noncognates were named more slowly with more divergent productions, while L1 cognates, L2 noncognates, and L2 cognates were named more quickly with more divergent productions. The *L2 accent and interest* factor, addressed in Chapter 5, was also involved in an interaction with naming language and cognate status. With higher levels of the *L2 accent and interest* predictor naming was faster in all conditions apart from for L2 noncognates. This factor was associated with variables related to L2 accent, in addition to language general variables. It was suggested that participants with higher factor scores might be more focused on pronouncing the L2 correctly, leading to slower production, but that the facilitatory effects of form and meaning overlap would make this less problematic for L2 cognates. The more language general aspects of the factor, on the other hand, might account for the overall relationship with faster naming. For the divergence measure, it is more difficult to interpret why naming was faster with higher values of the predictor, but not for L1 noncognates.

Cognate benefits are attributed to the cross-linguistic overlap in form and meaning. There are many examples of their facilitatory effect (e.g., Costa et al., 2000; Rosselli et al., 2014), but also some indications that cognates can inhibit production in language switching tasks (e.g., Broersma et al., 2016). More divergent productions might indicate that L1 and L2 speech sound representation are less similar, which could potentially interact with cognate effects and speech production in general. For instance, speakers with more divergent productions might require more control to keep L1 and L2 sound categories apart, which might slow down production, compared to those with less divergent productions who could be producing vowels in both languages with a similar quality. In language switching tasks, more divergent feedback might be easier to identify as L1 or L2 compared to less divergent productions. However, no evidence of such effects was found in the current study. In the picture naming task, more divergent /ʌ/-type productions were associated with faster naming, apart from for L1 noncognates. For the more acoustically similar /uː/-type vowels, the predictor did not interact with cognate status and the positive effects of more divergence were stronger for L1. The only divergence measure retained in the models of the switching task latencies was the /uː/-type Pillai score. The pattern of results was similar to that found in the picture naming task, faster naming in both languages with more divergence, but a stronger effect for L1. However, this two-way interaction between the divergence measure and language was no longer significant after controlling for the false discovery rate.

The /ʌ/-type measures were associated with bilingual profile factors related to early language development and more L2 exposure. It is possible that this measure is linked to better language proficiency in general, which for this group of bilinguals could indicate that the bilinguals with more divergent productions are more balanced bilinguals. However, none of the divergence measures were significantly related to the L1 or L2 proficiency factors. Second, if this was a clear effect one would expect to see similar results for both the BA and Pillai score which was not the case in this study. Overall, only limited effects were observed for the articulatory divergence measures in the picture naming and language switching tasks. The effects that were observed were in the direction of faster naming with more divergent productions, apart from for L1 noncognates in the picture naming task. It is still possible that some of these effects are false positives, even though the false discovery rate was controlled. On the other hand, clearer effects might have been observed using a different divergence measure or in a more targeted study.

As mentioned, previous studies have found evidence of both cognate inhibition and facilitation in switching tasks, suggesting that the phonological similarity of cognates might not always be beneficial, especially in language switching (Section 5.2.3). Depending on the level of activation at different levels of processing the result could be either facilitation or inhibition. This in turn could be affected by phonological similarity between cognates in languages or individuals. In the current study, where the divergence measure was based on vowels measured in both cognate and noncognate words, the only effect of cognate status was observed for L1 noncognates. A future study, with more items, might measure divergence separately for vowels produced in cognate and noncognate words. As discussed previously, there are still many open questions regarding the representations of words and sound structure and the link between perception and production. A recent study investigating prelexical and lexical perception and production of L2 French /u/ and /y/ by L1 speakers of either American or British English found evidence for a relationship between perception and production, but only when the tasks assessed the same level of processing (Melnik-Leroy et al., 2022). The measures considered in the current project do not measure the same levels of processing by design, but more targeted measures could have provided clearer results and simplified interpretations. For instance, by calculating divergence measures separately for cognates and noncognates or asking participants to specifically rate proficiency in pronouncing vowels.

When searching for divergence measures, two options, the Pillai score and Bhattacharyya's Affinity (BA), that seemed appropriate for capturing F1-F2 variation were selected. While one previous study reported finding similar results with the Pillai score and BA (Labov et al., 2016), this was true for less than half of the effects observed in the current study. Two L1-L2 vowel contrasts were selected for the divergence measures based on difficulty for proficient speakers, acoustic similarity, and known L1 substitutions of L2 vowels. In general, the more acoustically similar /uː/-type vowels were more overlapping than the /ʌ/-type vowels which often are replaced with different L1 vowels depending on the spelling of the English word. A future study might more systematically investigate influences on Norwegian productions of L2 /ʌ/, as mentioned, by including measures of the typical substitutions L1 /œ/ and L1 /ɔ/, as well as the more acoustically similar L1 /ɑ/. The current results indicate, though not strongly, that more fine-grained differences are more appropriately captured by the /uː/-type vowels. The divergence measures could also potentially be improved by including F3, as the L1 vowels considered in the current study are produced with lip rounding, unlike the target L2 vowels.

There are also some methodological factors relating to data collection and formant measurements that can be reconsidered in future studies. Speech materials were not normalised in the current study as each person was their own control, and normalisation might mask relevant linguistic variation. However, other studies comparing vowel productions within speaker have normalised Hz values prior to analysis (Mairano et al., 2019). While no striking differences between broad L1 dialect groups were detected in the current study (see Section 6.3.4), there might for instance be small differences between vowel productions in Norwegian dialects or coarticulation effects that were not detected. The speech materials were collected from responses in the picture naming and language switching experiments, where experimental items were selected to meet a set of criteria (see Sections 5.3.1 and 5.3.2). While this limited the influence of orthography, the environments surrounding the vowels were not controlled and there were a limited number of tokens per vowel. A future, more targeted study might benefit from a design similar to that of Kartushina and Frauenfelder (2014), for instance, where pictures of monosyllabic words were used to elicit responses, but participants only responded with the target vowel. Another option could be using data from spoken corpora if these are available in the language(s) of interest and provide sufficient information about the speakers for looking at individual differences.

Individual differences in articulation were quantified by measuring the articulatory divergence between similar, but not identical, L1 and L2 vowels. The assumed links between L1 and L2 phonetic categories, and the outcomes for speech production are largely based on the SLM(-r). Importantly, this model focuses on speech *learning*, and the speakers in the current study are relatively proficient L2 speakers. However, the model makes predictions about the formation of L2 phonetic categories, and especially the SLM-r highlights individual differences that not only affect the process of learning phonetic categories, but also the outcomes of the learning process.

The SLM-r, as well as available information about typical L1 Norwegian substitutions in the production of L2 English vowels, informed the selection of vowels in the current study. Representing individual differences in articulation using specific vowels can be seen as a first step in investigating the relationship between individual differences in representations of L1 and L2 sound structure, proficiency, bilingual profile, and language production. The observed relationships were generally not the same for the two selected vowel pairs. The variation seemed to partially reflect the individual differences in phonological processing, but also variables relating to language exposure, and potentially other factors that were not considered in the current study. Another approach might be using more general measures of divergence or variability in production, for instance using a larger set of vowels, or the whole vowel space, as in Kartushina and Frauenfelder (2014).

6.5.1 Conclusion

This chapter investigated how individual differences in articulation, quantified by four divergence measures, relate to self-reported and tested proficiency, bilingual profile measures and language production latencies in a picture naming and switching task. In general, there were many comparisons and few significant results. One clear observation is that different effects were found depending on which divergence measure (Pillai score or BA) and which vowel pair (/ʌ/-type or /uː/-type) was used to quantify individual differences in articulation. None of the articulatory divergence measures related to self-ratings of accent proficiency, but there were some significant relationships with the language test scores. The /uː/-type Pillai score was significantly related to L1 SNWR scores and marginally related to L2 spelling scores. The /uː/-type BA was also significantly related to L1 SNWR scores and the /ʌ/-type Pillai score was significantly related with the L2 spelling scores. This suggests that there is a link between the divergence measures and tested proficiency in domains relating to phonology, however these effects may be limited to specific aspects as there were no significant effects for elision scores in either language, L2 SNWR scores or L1 spelling scores.

The analysis found no relationship between /uː/-type divergence measures and the 10 factors extracted from the bilingual profile questionnaire. Both the Pillai score and BA for the /ʌ/-type vowels were associated with the factors *Language development*, *L2 exposure and mixing*, and *L1 informal exposure*. More divergent productions relate to reaching language development milestones at an earlier age and more L2 exposure and language mixing. The link between more divergent productions with more L2 exposure, especially communicative exposure, is partially supported by the significant relationship between the Pillai and *L2 interaction and mixing*, and the BA and *Informal learning*. However, these results are not clear as there is also an association between informal exposure to the L1 and more divergent productions. It is discussed whether individual differences in the production of the more acoustically similar /uː/-type vowels might be more related to specific phonological abilities, while individual differences in the production of /ʌ/-type vowels are more related to variables associated with general abilities and L2 exposure. However, there is only partial support for these interpretations.

Finally, there were only limited effects of articulatory divergence in the language production data. Each divergence measure was added as a predictor to separate models of the picture naming and language switching data. The predictors were only retained in three of the models. In the picture naming task, there was a two-way interaction between naming language and the /uː/-type Pillai score, and a three-way interaction between the /ʌ/-type BA, cognate status, and naming language. The /uː/-type Pillai score was retained in the model of the switching data, but there were no significant effects after controlling for the false discovery rate. The effects that were observed were in the direction of faster naming with more divergent production, apart from for L1 noncognates in the picture naming task. Overall, the significant effects that were observed were in the direction of more divergent production with higher language test scores and faster language production, however there were also many insignificant comparisons. Future studies would benefit from a more targeted design and some suggestions were made in the discussion. Individual differences in articulation may contribute to the description of bilinguals and account for bilingual linguistic behaviour, but the results from the current study do not allow for any clear conclusions.

7 General discussion

The aim of this thesis was to investigate the effects of individual differences in bilingual profile and articulatory divergence on bilingual spoken word production. Using the classic psycholinguistic paradigms of picture naming and language switching, previously documented effects of cognate status and naming language were also investigated in a new population of bilinguals. Several types of data were used, including a bilingual profile questionnaire, objective tests of language proficiency, phonetic measures of L1 and L2 divergence, and psycholinguistic experiments testing spoken word production and language switching.

In bilingual word production and language switching, general language proficiency differences have been shown to modulate naming behaviour (e.g., Costa & Santesteban, 2004). The motivation for focusing on individual differences in the domains of phonology and accent in the current study, and their effects on word production, came from observations of cognate effects in word production (e.g., Costa et al., 2000; Hoshino & Kroll, 2008). Cognate effects are partly attributed to the phonological similarity between cognate words (e.g., Hoshino & Kroll, 2008). The degree of phonological similarity between cognates does not only depend on the words and languages involved, but also individual differences in representations of sound structure. This raises the possibility that how speech sounds are represented and articulated could modulate cognate effects and cross-language competition, thereby influencing the speed of naming and language switching. There is also some evidence that acoustic characteristics of speech can be affected by cognate status both in word naming (Amengual, 2016a) and language switching (Goldrick et al., 2014).

The assessment of bilingual profile (including general L2 proficiency and the domains of phonology and accent), objective language tests, and measures of articulatory divergence involved several methods that were developed for this thesis or used in a different way compared to previous studies. This study also employed the more well-established psycholinguistic paradigms of simple picture naming and picture naming with language switching. The general discussion will first consider the effects of cognate status, bilingual profile predictors and articulatory divergence in bilingual word production. Then the relationships between the divergence measures, bilingual profile factors and language tests will be discussed.

7.1 Cognate effects and individual differences in bilingual word production

Cognates facilitated production in both the picture naming task and the language switching task, supporting findings of a benefit with form overlap in production. The *General L2 proficiency* and *L2 accent and interest* factors were added to the models of the picture naming and language switching data to assess whether aspects of bilingual profile modulate bilingual word production. The *General L2 proficiency* predictor did not influence naming latencies in either task, a finding which is in line with previous studies as the participants were a relatively proficient and homogeneous group of bilinguals.

In the picture naming task, higher values of the *L2 accent and interest* predictor were associated with faster naming, however a three-way interaction with cognate status and naming language indicated that this was not the case for L2 noncognates. In L2 therefore, the difference between cognate and noncognate latencies increased with higher values of the predictor (a larger cognate benefit), whereas in L1 this difference decreased with higher values of the predictor (a smaller cognate benefit). One possible explanation is that participants with higher scores on the *L2 accent and interest* factor might be more focused on pronouncing the L2 correctly which could lead to relatively slower production of L2 noncognates in particular. The general production benefit associated with cognates might mitigate this effect for L2 cognates. As naming in both L1 conditions was faster with higher values of the predictor, it was also considered that this factor might also reflect some language general aspects that are associated with faster production. The *L2 accent and interest* factor contains two positive loadings of language general variables, and correlated with L2 spelling and L1 elision scores, which might support this interpretation.

As an objective measure of L1-L2 articulatory differences, two measures of articulatory divergence, the Pillai score and Bhattacharyya's Affinity (BA), were calculated for two pairs of similar L1 and L2 vowels, the /uː/-type vowels (L1 /ʉː/ and L2 /uː/) and the /ʌ/-type vowels (L1 /œ/ and L2 /ʌ/). The influence of the four divergence measures on picture naming and switching latencies were assessed separately for each task and measure, giving a total of 8 models. The effects of these measures were subtle at best, and it is important to note that there were many insignificant comparisons, and that five of the predictors were not retained in the models of reaction times. In picture naming, more divergent productions measured by the /ʌ/-type BA were generally associated with faster naming, apart from for L1 noncognates. The results indicated that in L2 more divergent productions were associated with smaller latency differences between cognate and noncognate words (a smaller cognate benefit), while in L1 more divergent productions were associated with larger differences between cognate and noncognate words (a larger cognate benefit). This is the reverse of the pattern observed for the *L2 accent and interest* predictor discussed above and the reasons for this are unclear. It does however suggest that these predictors are measuring different things, we return to this issue in the following sections. With the /uː/-type Pillai score more divergence was again associated with faster naming. This effect was stronger for L1 than L2 but did not interact with cognate status. The presence of cognate effects in bilingual word production inspired the investigation into effects of individual differences in the domains of phonology and accent. The limited effects observed here do not suggest that cognate effects are strongly modulated by individual differences in these domains, at least not when assessed by the measures employed here.

In the more challenging switching task, cognates facilitated production in both L1 and L2, and the effect was stronger in L1. There were no significant effects of the *General L2 proficiency* and *L2 accent and interest* predictors, but they were each involved in a marginal three-way interaction. The /uː/-type Pillai score was retained in the model of the switching data, but there were no significant effects after controlling for the false discovery rate.

Cognate effects were uniformly beneficial in the current study. However, it has been argued that cognate effects, especially in language switching, may be facilitatory at the phonological level, but inhibitory at the lexical level. If more divergent productions relate to more divergent representations, there is a possibility that in switching, more divergent L1 and L2 productions provide less confusing feedback, compared to less divergent feedback which may be harder to identify as either L1 or L2. In simple picture naming, if there is an effect of this nature, one might expect more divergence, i.e., less similar representations, to be less beneficial to production than more overlapping productions. This is not only dependent on the vowels, of course, but other segments, suprasegmental aspects, and overlap at the conceptual level. As such, the effects of divergence may be very small if they are present at all. There may be several parallel processes working to produce either inhibition or facilitation which cannot be separated in the current study. This adds to the complexity of interpreting these results, and the only pattern that is relatively clear is that faster production in picture naming is associated with higher values of the *L2 accent and interest* predictor, as well as two of the four divergence measures.

Importantly, the limited effects of the *L2 accent and interest* and articulatory divergence predictors were found in the context of clear replications of several previously reported findings in picture naming and language switching, with a new population of bilinguals. In addition to cognate facilitation, this includes effects of naming language and trial type. In the simple picture naming task, L1 naming was faster than L2 naming, while this pattern was reversed in the switching task, suggesting the involvement of global inhibition of the L1 in the latter task. In the switching task, switch trials were slower and more error prone than stay trials. The participants in the current study were L1 dominant, but proficient in both their L1 and L2, making them relatively balanced bilinguals. The current study therefore also replicated findings of symmetric switch costs for more balanced bilinguals.

In addition to the limited effects of the linguistic predictors on word production, this study found some influence of individual differences in non-linguistic attentional control on word production. Better attentional control, indicated by flanker task performance, interacted with cognate status in the picture naming task. This was interpreted as an increased need for control when producing noncognate words compared to cognate words. In the switching task, which requires more control in general, the benefit was equally present across conditions. These results suggest that there is a benefit of better non-linguistic attentional control in speech production. The lack of interactions with naming language and picture naming session order in both tasks, as well as cognate status and trial type in the switching task, supports the involvement of additional control mechanisms in linguistic processing.

Finally, unpredicted, but interesting effects of language order in the simple picture naming sessions were observed on performance in both the picture naming sessions and the following language switching session. The simple picture naming tasks were either conducted in L1 in the first session and L2 in the second session or vice versa. The data from the picture naming sessions were pooled for the analysis. Language switching was always completed in the third and final session. The three sessions were all completed on separate days. The order of languages in the picture naming sessions affected behaviour both in the picture naming task and language switching task. In the picture naming task, both cognate and noncognate naming latencies were shorter for the more dominant L1 for participants who named pictures in L1 in the first session. However, when pictures were named in L2 in the first session, L1 naming was slower in general, but the cognate benefit remained robust. These findings were interpreted to suggest that long-term L1 inhibition could persist for several days.

In the switching task, there was an interaction between session, cognate status, and trial type. There was cognate facilitation in all conditions, and naming was generally faster in the language that was used in the second picture naming session. Cognates behaved similarly in the switching task, irrespective of what language they were first named in, while there was a benefit for noncognate stay trials in the most recently used language. Naming language was not involved in this interaction, and it was argued that this interaction could reflect a language non-specific recency benefit, in contrast to the effects observed in the simple picture naming task that were attributed to long term L1 inhibition. A similar pattern was found in the non-significant three-way interaction between naming language, cognate status, and trial type. One possible explanation is that sleep consolidation of picture-name associations formed on the first day (in either L1 or L2) leads to them being processed similarly to an L1, while picture-name associations formed on the second day (in either L2 or L1) are being processed more like an L2. While there are robust effects of naming language overall in the current study, the effects of session represent a potential confound that should be considered in the design of future studies. As discussed above, in relation to cognate effects in switching, there may be several parallel processes involved in working to produce either inhibition or facilitation. The interactions between these processes remain poorly understood, and future studies are required to learn more about how these processes interact.

7.2 The relationships among predictors and other measures

As discussed above, the motivation for focusing on individual differences in the domains of phonology and accent in the current study, and their effects on word production, came from observations of cognate effects in word production. Assessing individual differences in these domains is not straightforward. It is therefore of interest to investigate the relationships between measures of individual differences in the domains of phonology and accent, and how these relate to each other and other aspects of bilingual profile and objective tests of language proficiency. These findings are revisited below.

The relationship between self-ratings in the domains of phonology and accent and self-ratings of other areas of bilingual profile were investigated through factor analysis. Variables relating to phonology and accent generally grouped together and contributed to a clear accent factor (*L2 accent and interest*), while language proficiency in other domains generally grouped together in two factors called *General L2 proficiency* and *General L1 proficiency*. However, there was some overlap between factors and the variable specifically asking participants to rate their proficiency in L2 pronunciation was part of the *General L2 proficiency* factor, and not the *L2 accent and interest* factor. The findings suggest that the questions added to the questionnaire can contribute to assessing the domains of phonology and accent, or more specifically accent, but that this construct is not completely independent from general L2 proficiency and factors relating to language exposure.

The outcome of the factor analysis suggested some overlap between self-ratings in the domains of phonology and accent and other aspects of bilingual profile. This was also indicated when looking at the relationships between factor scores and language test scores. For instance, higher L2 spelling scores were related to higher factor scores for both the L1 and L2 general proficiency factors, *L2 accent and interest* and *L2 exposure and mixing*. For the phonological tests, higher L1 elision scores were associated with more *L2 exposure and mixing*, less *Self-instruction* and reaching language development milestones at an earlier age, in addition to the positive correlation with *L2 accent and interest*. Higher L2 elision scores were associated with higher *General L2 proficiency*, as well as *L2 exposure and mixing* and *L2 interaction and mixing.* In contrast, the L1 SNWR scores were not related to any of the factor scores, while higher L2 SNWR scores were associated with reaching language development milestones at an earlier age.

As previous studies have reported links between self-rated and tested proficiency, the relationships between domain specific self-ratings of proficiency (spelling, vocabulary, and pronunciation) and test scores were also examined. Self-ratings of spelling proficiency and spelling test scores were related in both languages, while self-ratings of vocabulary proficiency and vocabulary scores only correlated for English. SNWR scores did not relate to pronunciation proficiency in either language. Interestingly, self-ratings of pronunciation proficiency related to elision scores in both languages. L1 elision, but none of the other phonological tests, related to the *L2 accent and interest* factor. In general, the findings suggest that there is a relationship between self-ratings in the domains of phonology and accent and tested proficiency, but also that objective measures assessing individual differences in the domains of phonology and accent (SNWR and elision) appear to address different aspects of these skills. The pattern suggests that the two phonological tests measure different aspects of phonological memory which are related to different aspects of language proficiency and bilingual profile. The elision task, which involves manipulating and producing speech segments, may to a larger extent relate to a participant's current language behaviour, while the SNWR measures aspects of phonological processing and memory that may be more important in early language development. This is partially supported by the correlation between L2 SNWR scores and the *Language development* factor, however L1 SNWR scores did not relate to this factor while L1 elision scores did, so again there is some overlap.

Turning to the articulatory divergence measures, we again see different relationships with performance on the two phonological tests. More divergent productions of the /uː/-type vowels were related to higher L1 SNWR scores, but there were no other significant relationships with the phonological tests. The relationship between the /uː/-type divergence measures and the L1 SNWR task was discussed in Chapter 6 with reference to the SLM-r. Herein, it is hypothesised that the precision of L1 categories at the time of L2 speech learning onset can influence the likelihood of forming categories for L2 speech sounds. It was argued that higher scores on the L1 SNWR task could reflect more precise L1 categories, which in turn increases the likelihood that the vowels are produced with different, more divergent, qualities. As the /uː/-type vowels are more acoustically similar than the /ʌ/-type vowels, it was suggested that divergent production of /uː/-type vowels might be more dependent on fine-grained individual differences in L1 category precision or auditory acuity compared to the /ʌ/-type vowels.

More divergent productions of the /ʌ/-type vowels, quantified by the Pillai score but not the BA, were related to higher L2 spelling scores. Higher L2 spelling scores were also positively correlated with the factors *General L1 proficiency*, *General L2 proficiency*, *L2 accent and interest* and *L2 exposure and mixing*. Both /ʌ/-type measures were associated with the *Language development* factor, that is, more divergent productions are associated with reaching language development milestones at an earlier age. This factor was also positively correlated to L2 SNWR scores and L1 elision scores. Norwegian L1 speakers tend to substitute L2 English /ʌ/ with vowel qualities similar to L1 /œ/ or /ɔ/, which has been attributed to influences from spelling, as speakers have access to the more acoustically similar L1 sound /ɑ/. In the current study there was more overlap between L1 /œ/ and L2 /ʌ/, than L1 /ɔ/ and L2 /ʌ/, so the former pair was selected for the divergence measure. These effects are consistent with strong relationships between phoneme and grapheme representations and could suggest that more divergent productions of the /ʌ/-type vowels are related to a more language general aptitude in the phonological domain, compared to the /uː/-type vowels. It should also be noted that there was marginal relationship between the /uː/-type Pillai score and L2 spelling, so this is not exclusive to the /ʌ/-type vowels. The interpretation of the links between the *Language development* factor, test scores and /ʌ/-type divergence would be more compelling if both L1 and L2 scores for the measures involved correlated with the factors.

Both /ʌ/-type measures were also associated with the factors *L2 exposure and mixing*, and *L1 informal exposure*. In addition, the Pillai score was related to *L2 interaction and mixing* and the BA was related to *Informal learning*. There were no significant relationships between the /uː/-type measures and the bilingual profile factors. The SLM(-r) is explicit about the importance of language exposure, in addition to the precision of L1 categories, for L2 speech learning. This could suggest that individual differences in the production of /ʌ/-type vowels are more related to variables associated with L2 exposure and a language general aptitude in the phonological domain, while the more acoustically similar /uː/-type vowels, as suggested above, are more dependent on fine-grained individual differences in L1 category precision or auditory acuity. However, there was only partial support for these interpretations of the difference between the /uː/-type and /ʌ/-type vowels.

Critically, we found no links between the divergence measures and self-ratings of pronunciation proficiency. As discussed in Chapter 6, when participants were asked if they had noticed any specific L2 speech sounds that were difficult to pronounce for themselves and others, most of them responded with consonants. It is therefore possible that the results would have been different if they specifically had been asked to rate their own proficiency in pronouncing vowels. The current study did not compare the articulatory divergence measures with native ratings or other evaluations of accentedness. One study measuring acoustic distances in tense-lax pairs of L2 vowels, found that these measures were correlated with ratings of comprehensibility and nativelikeness to some extent (Mairano et al., 2019), and these links could be investigated further. It has also been argued that "cross-language dissimilarity must be addressed perceptually rather than acoustically" (Flege & Bohn, 2021, p. 33). However, perceptual measures have their drawbacks as they are also inevitably influenced by individual differences in the perceivers. Although a number of studies have shown a relationship between perception and production accuracy (e.g., W. Baker & Trofimovich, 2006; Flege, 1999; Kim & Clayards, 2019; but see, Kartushina & Frauenfelder, 2014; Peperkamp & Bouchon, 2011), future research still is required that combines articulatory divergence measures and perceptual tests to develop an understanding of how perception and production is related, and to ascertain which aspects may be the most informative for assessing individual difference effects.

7.3 Assessing individual differences in phonology and accent - conclusions

This thesis constitutes a first attempt to develop methods to assess individual differences in the domains of phonology and accent and to relate them to bilingual language production. Significant relationships were observed between bilingual profile factors, objective language tests, articulatory divergence measures and self-rated proficiency. However, the resulting pattern is not always easy to interpret. Defining and measuring proficiency in the domains of phonology and accent remains complex. The bilingual questionnaire contained additional questions relating to the domains of phonology and accent, and several of the objective tests employed here were developed for this thesis. While this makes it difficult to compare the findings with previous studies, the broad approach employed in this thesis allows for a comprehensive description of a specific group of bilinguals in a psycholinguistic context.

In the current analysis, bilingual profile predictors were limited to two theoretically motivated factors, *L2 accent and interest* and *General L2 proficiency*. While the *General L2 proficiency* factor was clearly related to tested L2 proficiency, the relationships between the *L2 accent and interest* and other measures meant to address the domains of phonology and accent were less clear. The factor analysis of the questionnaire responses yielded 10 factors, where the five first, including *L2 accent and interest*, were generally the most reliable. The *L2 accent and interest* factor was related to L2 spelling and L1 elision, but none of the other language test scores. Questions grouping together in this variable included finding L2 accent important, interest and effort in improving L2 accent, ability to imitate accents and paying attention to pronunciation. These questions were intended to target abilities in the phonological domain, as it is not possible to ask about phonology directly. The results suggested a degree of success in this endeavour. In this relatively proficient group of bilinguals, there was some evidence of individual differences in the domains of phonology and accent, assessed by the *L2 accent and interest* factor, affecting word production behaviour, but the effects were not very strong. The general tendency was faster production with higher *L2 accent and interest* factor scores. This factor combines a variety of abilities, and it remains unclear which are the key predictors of speed and control in L2 word production.

Individual differences were also observed in the articulatory divergence measures. The divergence measures did not relate to the *L2 accent and interest* factor, nor *General L2 proficiency* and *General L1 proficiency*, suggesting that individual differences in articulatory divergence are not related to general language proficiency. There was however some overlap with the divergence measures and two of the other more reliable factors, *L2 exposure and mixing* and *Language development*. The SLM(-r) is explicit about the importance of communicative language exposure (e.g., Flege & Bohn, 2021, p. 32) for L2 speech learning. Moreover, there is evidence of language exposure effects in switching (e.g., Bonfieni et al., 2019). A detailed investigation of the role of language exposure in L2 accent development and its role in bilingual spoken word production would be an interesting avenue for future research.

The current study focused on to what extent divergence in articulation affected bilingual word production. Individual differences in articulation may contribute to the description of bilinguals and accounting for bilingual linguistic behaviour, but the results from the current study do not present a simple picture. As discussed, future studies could benefit from more targeted designs and some adjustments to the tests used, including standardised tests if they are available in the languages of interest. The vowel contrasts selected for this thesis are specific to Norwegian-English bilinguals. Another option is using a measure of overall divergence or variability in production, for instance using a larger set of vowels, or the whole vowel space, as in Kartushina and Frauenfelder (2014). The vowel pairs selected for the articulatory divergence measures were motivated by the SLM, and available information about typical L1 Norwegian substitutions in the production of L2 English vowels. The observed relationships were generally not the same for the two selected vowel pairs. Possibly due to the different degree of overlap, the individual difference measures seemed to be related to different aspects of bilingual profile and tested proficiency. As the current thesis looked at cognate effects, an interesting next step could be to compare divergence measures taken from cognate and noncognate productions. While this is not ideal with respect to controlling context surrounding vowels of interest, employing heuristics for identifying stable portions of vowels seemed to work well with the semi-automatic approach for measuring formants used in this thesis. A future study with broader measures of divergence and more targeted measures of phonological abilities might be able to expand on some of the weak effects observed in the current study. An interesting focus might be the use of SNWR performance as a predictor. While elision was more related to factors and pronunciation proficiency, SNWR was more strongly related to the articulatory divergence measures.

7.4 Concluding remarks

This project constitutes a first step in the investigation of individual differences in bilingual profile, particularly relating to the domains of phonology and accent, and articulatory divergence on bilingual word production. Several previously reported findings in picture naming and language switching were replicated with a new population of bilinguals, including the facilitatory effects of cognates in bilingual word production. In picture naming, but not language switching, there were some indications of individual differences in the domains of phonology and accent affecting word production. The results suggested faster speech production was associated with higher scores on self-reported variables relating to L2 accent, and to some extent more articulatory divergent productions of similar, but not identical, L1 and L2 vowels. This research highlights the benefits and challenges of using an individual differences approach in the study of bilingual spoken word production. The results demonstrate a role for individual differences in the domains of phonology and accent, but for clearer answers the measures used need to be refined. These findings will inform the development of future studies aiming to assess individual differences in the domains of phonology and accent in bilingual language production.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. http://dx.doi.org/10.1017/S027226310808073X

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 207–245. https://doi.org/10.3102/0034654310368803

Aliaga-Garcia, C., Mora, J. C., & Cerviño-Povedano, E. (2010). Phonological short-term memory and L2 speech learning in adulthood. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Poznań Studies in Contemporary Linguistics* (pp. 19–24).

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, *15*(3), 517–530. https://doi.org/10.1017/S1366728911000460

Amengual, M. (2016a). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00617

Amengual, M. (2016b). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism*, *20*(2), 133–152. https://doi.org/10.1177/1367006914544988

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Anthony, J. J. D. R., & Blumenfeld, H. K. (2019). Language dominance predicts cognate effects and inhibitory control in young adult bilinguals. *Bilingualism: Language and Cognition*, *22*(5), 1068–1084. https://doi.org/10.1017/S1366728918001013

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*(4), 640–653. https://doi.org/10.1016/j.wocn.2010.09.005

Audacity Team. (2019). *Audacity®* (2.1.3). https://audacityteam.org/

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for 'Grid' Graphics* (2.3). https://CRAN.R-project.org/package=gridExtra

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Baker, C. (2001). *Foundations of bilingual education and bilingualism* (3rd ed.). Multilingual Matters.

Baker, W., & Trofimovich, P. (2006). *Perceptual paths to accurate production of L2 vowels: The role of individual differences*. *44*(3), 231–250. https://doi.org/10.1515/IRAL.2006.010

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barreda, S. (2021). Perceptual validation of vowel normalization methods for variationist research. *Language Variation and Change*, *33*(1), 27–53. https://doi.org/10.1017/S0954394521000016

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*(2), 77–85. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bergmann, C., Nota, A., Sprenger, S. A., & Schmid, M. S. (2016). L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics*, *58*, 71–86. https://doi.org/10.1016/j.wocn.2016.07.001

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in Factor Analysis. *Educational and Psychological Measurement*, *65*(5), 676–696. https://doi.org/10.1177/0013164404272507

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2013). From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition*, *127*(3), 287–306. https://doi.org/10.1016/j.cognition.2013.02.005

Best, C. T. (1994a). Learning to perceive the sound pattern of English. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research* (Vol. 8, pp. 217–304). Ablex publishers.

Best, C. T. (1994b). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). MIT Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). John Benjamins Publishing Company.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, *35*, 99–110.
Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001
Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *19*(1), 69–85. https://doi.org/10.1017/S0142716400010584
Bion, R. A. H., Escudero, P., Rauber, A. S., & Baptista, B. O. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. *Proc. Interspeech 2006*, paper 1270-Tue3CaP.3. https://doi.org/10.21437/Interspeech.2006-411
Bloomfield, L. (1933). *Language*. Holt.
Boersma, P., & Weenik, D. (2021). *Praat: Doing phonetics by computer* (6.1.51). http://www.praat.org
Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica*, *193*, 160–170. https://doi.org/10.1016/j.actpsy.2018.11.004
Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276
Brato, T. (2016). *TB-Track vowels* (2.1). https://www.uni-regensburg.de/language-literature-culture/english-linguistics/staff/brato/praat-scripts/index.html
Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010
Broersma, M., Carter, D., & Acheson, D. J. (2016). Cognate costs in bilingual speech production: Evidence from language switching. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01461
Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology* (pp. 341–376). Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.019
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977
Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press.
Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2012). Qualitative differences between bilingual language control and executive control:

Evidence from task-switching. *Frontiers in Psychology*, *2*, 399. https://doi.org/10.3389/fpsyg.2011.00399
Calenge, C. (2006). The package adehabitat for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling*, *197*, 516–519. https://doi.org/10.1016/j.ecolmodel.2006.03.017
Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177–208. https://doi.org/10.1080/026432997381664
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10
Cerviño-Povedano, E., & Mora, J. C. (2015). Spanish EFL learners' categorization of /iː-ɪ/ and phonological short-term memory. *Procedia - Social and Behavioral Sciences*, *173*, 18–23. https://doi.org/10.1016/j.sbspro.2015.02.024
Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*, 192–208. https://doi.org/10.1016/j.brainres.2007.01.137
Christopoulos, D. T. (2022). *inflection: Finds the Inflection Point of a Curve* (1.3.6). https://CRAN.R-project.org/package=inflection
Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge University Press.
Colomé, À., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 96–109. https://doi.org/10.1037/a0017677
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., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–149. https://doi.org/10.1016/j.cognition.2009.08.001
Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59–86. https://doi.org/10.1016/j.cognition.2006.12.013
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., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, *94*(1), 94–103. https://doi.org/10.1016/j.bandl.2004.12.002
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. https://doi.org/10.1007/BF02310555
de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, *9*(3), 33. https://doi.org/10.3390/bs9030033
De Groot, A. M. B., & Christoffels, I. K. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition*, *9*(2), 189–201. https://doi.org/10.1017/S1366728906002537
De Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology Section A*, *53*(2), 397–428. https://doi.org/10.1080/713755891
De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–91. https://doi.org/10.1017/S0142716412000069
Declerck, M., Grainger, J., Koch, I., & Philipp, A. M. (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language*, *95*, 138–145. https://doi.org/10.1016/j.jml.2017.03.005
Declerck, M., Koch, I., & Philipp, A. M. (2012). Digits vs. Pictures: The influence of stimulus type on language switching. *Bilingualism: Language and Cognition*, *15*(4), 896–904. Health Research Premium Collection. https://doi.org/10.1017/S1366728912000193
Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, *22*(6), 1630–1645. https://doi.org/10.3758/s13423-015-0836-1
Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioral Sciences*, *21*(1), 31–46. https://doi.org/10.1177/0739986399211003
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283
Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*(2), 124–142. https://doi.org/10.1016/0749-596X(88)90070-8
Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, *23*(4), 517–542. https://doi.org/10.1016/S0364-0213(99)00014-2
Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International*

Phonetic Association, *27*(1–2), 47–55. https://doi.org/10.1017/S0025100300005417
Diebold, A. R. (1961). Incipient bilingualism. *Language*, *37*(1), 97–112. https://doi.org/10.2307/411253
Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (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., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284–301. https://doi.org/10.1016/j.jml.2009.12.003
Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012
Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System*, *36*(4), 506–516. https://doi.org/10.1016/j.system.2008.03.003
Dufva, M., & Voeten, M. J. M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics*, *20*(3), 329–348. https://doi.org/10.1017/S014271649900301X
Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261
Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663–679. https://doi.org/10.1037/0278-7393.33.4.663
Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Netherlands National Graduate School of Linguistics. https://www.lotpublications.nl/linguistic-perception-and-second-language-acquisition-linguistic-perception-and-second-language-acquisition-explaining-the-attainment-of-optimal-phonological-categorization
Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), 452–465. https://doi.org/10.1016/j.wocn.2009.07.006
Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

European Commission. (2012). *Europeans and their languages*. https://op.europa.eu/en/publication-detail/-/publication/f551bd64-8615-4781-9be1-c592217dad83

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory Factor Analysis*. Oxford University Press.

Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, *6*(5). https://doi.org/10.1186/1744-9081-6-5

Field, A. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll* (4th ed.). SAGE.

Finch, W. H. (2020). *Exploratory factor analysis*. SAGE Publications, Inc. https://doi.org/10.4135/9781544339900

Fink, A., & Goldrick, M. (2015). Pervasive benefits of preparation in language switching. *Psychonomic Bulletin & Review*, *22*(3), 808–814. https://doi.org/10.3758/s13423-014-0739-6

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E. (1999). The relation between L2 production and perception. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetics Sciences* (pp. 1273–1276).

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353–382). De Gruyter Mouton.

Flege, J. E. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*, *21*(5), 919–920. https://doi.org/10.1017/S136672891800010X

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speeech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge University Press.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, *91*(1), 370–389. https://doi.org/10.1121/1.402780

Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*(4), 567–598. https://doi.org/10.1017/S0142716402004046

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. https://doi.org/10.1121/1.413041

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78–104. https://doi.org/10.1006/jmla.1999.2638

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., & Hagoort, P. (2017). Individual variability as a window on production-perception interactions in speech motor control. *The Journal of the Acoustical Society of America*, *142*(4), 2007–2018. https://doi.org/10.1121/1.5006899

Franken, M. K., McQueen, J. M., Hagoort, P., & Acheson, D. J. (2015). Assessing the link between speech perception and production through individual differences. In M. Wolters, J. Livingstone, & B. Beattie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. https://hdl.handle.net/2066/150165

Fricke, M., Zirnstein, M., Navarro-Torres, C., & Kroll, J. F. (2019). Bilingualism reveals fundamental variation in language processing. *Bilingualism: Language and Cognition*, *22*(1), 200–207. https://doi.org/10.1017/S1366728918000482

Fridland, V., & Kendall, T. (2012). Exploring the relationship between production and perception in the mid front vowels of U.S. English. *Lingua*, *122*(7), 779–793. https://doi.org/10.1016/j.lingua.2011.12.007

Fridriksson, J., Moser, D., Ryalls, J., Bonilha, L., Rorden, C., & Baylis, G. (2009). Modulation of frontal lobe speech areas associated with the production and perception of speech movements. *Journal of Speech, Language, and Hearing Research*, *52*(3), 812–819. https://doi.org/10.1044/1092-4388(2008/06-0197)

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 1–30. https://doi.org/10.1080/02724980042000002

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*(5), 887–898. https://doi.org/10.1037/0012-1649.28.5.887

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, *25*(4), 1031–1036. https://doi.org/10.1177/0956797613520014

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., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1122–1139. https://doi.org/10.1037/0278-7393.23.5.1122

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

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science*, *22*(9), 1155–1164. https://doi.org/10.1177/0956797611417002

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

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

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

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275. https://doi.org/10.1037/0033-2909.103.2.265

Guevara, E. R. (2010). NoWaC: a large web-based corpus for Norwegian. *Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop*, 1–7.

Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00278

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*(2), 283–294. https://doi.org/10.1017/S1366728919000026

Hall-Lew, L. (2010). Improved representation of variance in measures of vowel merger. *Proceedings of Meetings on Acoustics*, *9*. https://doi.org/10.1121/1.3460625

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

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Heikoop, K. W., Declerck, M., Los, S. A., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language

switching. *Bilingualism: Language and Cognition*, *19*(5), 921–927. https://doi.org/10.1017/S1366728916000456
Hepper, P. G., Scott, D., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*(3), 147–153. https://doi.org/10.1080/02646839308403210
Hernandez, A. E., & Meschyan, G. (2006). Executive function is necessary to enhance lexical processing in a less proficient L2: Evidence from fMRI during picture naming. *Bilingualism: Language and Cognition*, *9*(2), 177–188. https://doi.org/10.1017/S1366728906002525
Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131–138. https://doi.org/10.1016/S1364-6613(00)01463-7
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447
Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, *106*(1), 501–511. https://doi.org/10.1016/j.cognition.2007.02.001
Hulstijn, J. H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, *15*(2), 422–433. http://dx.doi.org/10.1017/S1366728911000678
Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842–2854. https://doi.org/10.1121/1.2783198
Jevtović, M., Duñabeitia, J. A., & de Bruin, A. (2020). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, *23*(2), 401–413. https://doi.org/10.1017/S1366728919000191
Jilka, M. (2009). Talent and proficiency in language. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and brain activity* (Vol. 1, pp. 1–16). De Gruyter Mouton. https://doi.org/10.1515/9783110215496
Johnson, D. E. (2015, October). *Quantifying overlap with Bhattacharyya's affinity*. 44th annual meeting on New Ways of Analyzing Variation (NWAV 44), Toronto. https://danielezrajohnson.shinyapps.io/nwav_44
Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116
Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401–415. https://doi.org/10.1007/BF02291817

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, Article 1246. https://doi.org/10.3389/fpsyg.2014.01246

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics*, *57*, 21–39. http://dx.doi.org/10.1016/j.wocn.2016.05.001

Kelley, M. C., & Tucker, B. V. (2020). A comparison of four vowel overlap measures. *The Journal of the Acoustical Society of America*, *147*(1), 137–145. https://doi.org/10.1121/10.0000494

Kendall, T., & Fridland, V. (2012). Variation in perception and production of mid front vowels in the U.S. Southern Vowel Shift. *Journal of Phonetics*, *40*(2), 289–306. https://doi.org/10.1016/j.wocn.2011.12.002

Khateb, A., Shamshoum, R., & Prior, A. (2017). Modulation of language switching by cue timing: Implications for models of bilingual language control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1239–1253. https://doi.org/10.1037/xlm0000382

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience*, *34*(6), 769–786. https://doi.org/10.1080/23273798.2019.1582787

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford University Press.

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

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 289–307). Oxford University Press.

Labov, W., Fisher, S., Gylfadottír, D., Henderson, A., & Sneller, B. (2016). Competing systems in Philadelphia phonology. *Language Variation and Change*, *28*(3), 273–305. https://doi.org/10.1017/S0954394516000132

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, *89*(1), 30–65. https://doi.org/10.1353/lan.2013.0015

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 637–675). De Gruyter Mouton.
Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002
Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language*, *92*, 217–233. https://doi.org/10.1016/j.jml.2016.06.010
Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.
Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232. https://doi.org/10.1016/S1364-6613(99)01319-4
Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*(23), 13464–13471. https://doi.org/10.1073/pnas.231459498
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75. https://doi.org/10.1017/S0140525X99001776
Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*(1), 239–269. https://doi.org/10.1016/0010-0277(94)90030-2
Li, C., & Gollan, T. H. (2018). Cognates facilitate switches and then confusion: Contrasting effects of cascade versus feedback on language selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6), 974–991. https://doi.org/10.1037/xlm0000497
Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210. https://doi.org/10.3758/BF03192770
Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673–680. https://doi.org/10.1017/S1366728913000606
Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language history questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 1–7. https://doi.org/10.1017/S1366728918001153
Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6
Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*(3), 251–285. https://doi.org/10.1017/S0272263100009165

Love, T., Maas, E., & Swinney, D. (2003). The influence of language exposure on lexical and syntactic language processing. *Experimental Psychology*, *50*(3), 204–216. https://doi.org/10.1026//1617-3169.50.3.204

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Luk, G., de Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*(4), 588–595. http://dx.doi.org/10.1017/S1366728911000010

Ma, F., Li, S., & Guo, T. (2016). Reactive and proactive control in bilingual word production: An investigation of influential factors. *Journal of Memory and Language*, *86*, 35–59. https://doi.org/10.1016/j.jml.2015.08.004

Ma, W., & Winke, P. (2019). Self-assessment: How reliable is it in assessing oral proficiency over time? *Foreign Language Annals*, *52*(1), 66–86. https://doi.org/10.1111/flan.12379

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, *47*(2), 265–287. https://doi.org/10.1111/0023-8333.81997008

MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, *58*(1–2), 103–125. https://doi.org/10.1159/000028490

Mairano, P., Bouzon, C., Capliez, M., & de Iacovo, V. (2019). Acoustic distances, Pillai scores and LDA classification scores as metrics of L2 comprehensibility and nativelikeness. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*.

Mangiafico, S. S. (2016). *Summary and analysis of extension program evaluation in R, version 1.18.1.* rcompanion.org/documents/RHandbookProgramEvaluation.pdf

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)

Marian, V., & Shook, A. (2012). The cognitive benefits of being bilingual. *Cerebrum: The Dana Forum on Brain Science*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3583091/

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(1), 81–93. https://doi.org/10.1017/S1366728907003227

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

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(3), 279–297. https://doi.org/10.1017/S1366728909990022

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using kaldi. *Proc. Interspeech 2017*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research*, *38*(3), 581–605. https://doi.org/10.1177/0267658320988061

Meuter, R. F. I., & 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

Minkova, D. (2014). *A historical phonology of English*. Edinburgh University Press.

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

Mora, J. C., & Mora-Plaza, I. (2019). Contributions of cognitive attention control to L2 speech learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 477–499). Aarhus university.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 555–563. https://doi.org/10.1037//0278-7393.28.3.555

Nakayama, M., Verdonschot, R. G., Sears, C. R., & Lupker, S. J. (2014). The masked cognate translation priming effect for different-script bilinguals is modulated by the phonological similarity of cognate words: Further support for the phonological account. *Journal of Cognitive Psychology*, *26*(7), 714–724. https://doi.org/10.1080/20445911.2014.953167

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*(3), 359–377. https://doi.org/10.1016/j.jml.2005.05.001

Nilsen, T. S. (2010). *English pronunciation and intonation: British, American and world Englishes* (3rd ed.). Universitetsforlaget.

Nilsen, T. S., & Rugesæter, K. N. (2015). *English phonetics for teachers* (3rd ed.). Fagbokforlaget.

Nycz, J., & Hall-Lew, L. (2014). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*, *20*, 1–19. https://doi.org/10.1121/1.4894063

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, *27*(3), 377–402. http://dx.doi.org/10.1017/S0142716406060322

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(4), 557–581. https://doi.org/10.1017/S027226310707043X

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, *41*(6), 407–420. https://doi.org/10.1016/j.wocn.2013.07.005

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, *9*, Article 6. https://doi.org/10.7275/qf69-7k43

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*(2), 89–112. https://doi.org/10.1080/20445911.2016.1248436

Papazian, E., & Helleland, B. (2005). *Norsk talemål: Lokal og sosial variasjon*. Høyskoleforlaget.

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. *Proc. Interspeech 2011*, 161–164. https://doi.org/10.21437/Interspeech.2011-72

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (Vol. 45, pp. 137–157). John Benjamins Publishing Company.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154. https://doi.org/10.1177/00238309030460020501

Poarch, G. J., & van Hell, J. G. (2012). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *111*(3), 419–438. https://doi.org/10.1016/j.jecp.2011.09.008

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlíček, P., Qian, Y., Schwarz, P., Silovský, J., Stemmer, G., & Veselý, K. (2011). The Kaldi speech recognition toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.

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. http://dx.doi.org/10.1017/S1355617711000580

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(2), 253–262. https://doi.org/10.1017/S1366728909990526

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (4.1.3). R Foundation for Statistical Computing. https://www.R-project.org/

Raîche, G., & Magis, D. (2022). *nFactors: Parallel Analysis and Other Non Graphical Solutions to the Cattell Scree Test* (2.4.1.1). https://CRAN.R-project.org/package=nFactors

Raîche, G., Walls, T. A., Magis, D., Riopel, M., & Blais, J.-G. (2013). Non-Graphical solutions for Cattell's scree test. *Methodology*, *9*(1), 23–29. https://doi.org/10.1027/1614-2241/a000051

Reddy, S., & Stanford, J. (2015). A web application for automated dialect analysis. *Proceedings of NAACL-HLT 2015*.

Reetz, H. (2022). https://github.com/HenningReetz/Praat-scripts

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research* (2.3.3). https://CRAN.R-project.org/package=psych

Roberts, A. C., Wetterlin, A., & Lahiri, A. (2013). Aligning mispronounced words to meaning: Evidence from ERP and reaction time studies. *The Mental Lexicon*, *8*(2), 140–163. https://doi.org/10.1075/ml.8.2.02rob

Roelofs, A., & Verhoef, K. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition*, *9*(2), 167–176. https://doi.org/10.1017/S1366728906002513

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, *15*(1), 1–20. https://doi.org/10.1177/026553229801500101

Rosselli, M., Ardila, A., Jurado, M. B., & Salvatierra, J. L. (2014). Cognate facilitation effect in balanced and non-balanced Spanish–English bilinguals using the Boston Naming Test. *International Journal of Bilingualism*, *18*(6), 649–662. https://doi.org/10.1177/1367006912466313

Saiegh-Haddad, E. (2019). What is phonological awareness in L2? *Journal of Neurolinguistics*, *50*, 17–27. https://doi.org/10.1016/j.jneuroling.2017.11.001

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, *67*(3), 665–693. https://doi.org/10.1111/lang.12244

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439–462. https://doi.org/sant

Santesteban, M., & Costa, A. (2016). Are cognate words 'special'? In J. W. Schwieter (Ed.), *Cognitive Control and Consequences of Multilingualism* (pp. 97–125). John Benjamins Publishing Company.

Sasisekaran, J., & Byrd, C. (2013). Nonword repetition and phoneme elision skills in school-age children who do and do not stutter. *International Journal of Language & Communication Disorders*, *48*(6), 625–639. https://doi.org/10.1111/1460-6984.12035

Schepens, J., Dijkstra, T., Grootjen, F., & van Heuven, W. J. B. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLOS ONE*, *8*(5), e63006. https://doi.org/10.1371/journal.pone.0063006

Schmid, M. S., & Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing*, *31*(3), 367–388. https://doi.org/10.1177/0265532214526175

Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology*, *24*(5), 591–601. https://doi.org/10.1080/20445911.2012.669367

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, *164*, 77–105. https://doi.org/10.1016/j.bandl.2016.10.004

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

Stanley, J. (2019). *A tutorial in measuring vowel overlap in R*. https://joeystanley.com/blog/a-tutorial-in-calculating-vowel-overlap

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, *111*(4), 1872–1891. https://doi.org/10.1121/1.1458026

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America*, *109*(4), 1691–1704. https://doi.org/10.1121/1.1353594

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson/Allyn and Bacon.

Thorin, J., Sadakata, M., Desain, P., & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *The Journal of the Acoustical Society of America*, *144*(1), 92–103. https://doi.org/10.1121/1.5044415

Timmer, K., Christoffels, I. K., & Costa, A. (2019). On the flexibility of bilingual language control: The effect of language context. *Bilingualism: Language and Cognition*, *22*(3), 555–568. https://doi.org/10.1017/S1366728918000329

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, *22*(3), 516–536. https://doi.org/10.1017/S1366728918000421

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, *19*(1), 122–140. https://doi.org/10.1017/S1366728914000832

Vaden, K. I., Halphin, H. R., & Hickok, G. S. (2009). *Irvine Phonotactic Online Dictionary, Version 2.0. [Datafile]*. IPhOD. http://www.iphod.com

van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, *62*, 148–171. https://doi.org/10.1111/j.1467-9922.2012.00710.x

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*. http://dx.doi.org/10.3389/fpsyg.2015.01000

van Son, R. J. J. H., & Pols, L. C. W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *The Journal of the Acoustical Society of America*, *88*(4), 1683–1693. https://doi.org/10.1121/1.400243

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, *110*(1), 84–99. https://doi.org/10.1016/j.cognition.2008.10.013

Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, *19*(1), 181–190. https://doi.org/10.1017/S1366728914000352

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*(1), 1–27. https://doi.org/10.1111/bjop.12207

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408. https://doi.org/10.1006/jmla.1998.2618

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, *145*(2), 246–258. https://doi.org/10.1037/xge0000120

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wharton, G. (2000). Language learning strategy use of bilingual foreign language learners in Singapore. *Language Learning*, *50*(2), 203–243. https://doi.org/10.1111/0023-8333.00117

Wheeldon, L. R., & Konopka, A. E. (2018). Wheeldon, L. R., & Konopka, A. (2018). Spoken word production: Representation, retrieval, and integration. The Oxford Handbook of Psycholinguistics. In S.-A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford Handbook of Psycholinguistics* (2nd ed.). Oxford University Press.

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

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. https://CRAN.R-project.org/package=cowplot

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, *8*(3). https://doi.org/10.33151/ajp.8.3.93

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *ArXiv*, *1308.5499*. https://doi.org/10.48550/arXiv.1308.5499

Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1579–1586. https://doi.org/10.1037/xlm0000107

Wu, E.-C. (2011). Similar vowels in L1/L2 production: Confused or discerned in early L2 English learners with different amount of exposure. *Proc. Interspeech 2011*, 1141–1144. https://doi.org/10.21437/Interspeech.2011-339

Wu, Y. J., Cristino, F., Leek, C., & Thierry, G. (2013). Non-selective lexical access in bilinguals is spontaneous and independent of input monitoring: Evidence from eye tracking. *Cognition*, *129*(2), 418–425. https://doi.org/10.1016/j.cognition.2013.08.005

Yang, J., & Fox, R. A. (2017). L1–L2 interactions of vowel systems in young bilingual Mandarin-English children. *Journal of Phonetics*, *65*, 60–76. https://doi.org/10.1016/j.wocn.2017.06.002

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition*, *3*(2), 131–149. https://doi.org/10.1017/S1366728900000225

Zhang, K., & Peng, G. (2017). The relationship between the perception and production of non-native tones. *Proc. Interspeech 2017*, 1799–1803. https://doi.org/10.21437/Interspeech.2017-714

Zheng, X., Roelofs, A., & Lemhöfer, K. (2018). Language selection errors in switching: Language priming or cognitive control? *Language, Cognition and Neuroscience*, *33*(2), 139–147. https://doi.org/10.1080/23273798.2017.1363401

Zhou, B., & Krott, A. (2018). Bilingualism enhances attentional control in non-verbal conflict tasks – evidence from ex-Gaussian analyses. *Bilingualism: Language and Cognition*, *21*(1), 162–180. https://doi.org/10.1017/S1366728916000869

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–422. https://doi.org/10.1037/0033-2909.99.3.432

Appendices

Appendix A: Information sheet and consent form

Project ELL2

PARTICIPANT INFORMATION SHEET AND CONSENT FORM

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

REPRESENTATION AND PROCESSING OF ENGLISH AS A SECOND LANGUAGE

We are looking for native speakers of Norwegian to take part in a language study investigating the processing of English as a second language.
In order to participate in this study you need to be a **native speaker** of Norwegian with no other home languages (excluding perhaps English) and have a reasonable proficiency in English as your second language. You should have normal or corrected-to-normal vision (including colour vision) and hearing and have no diagnosed language impairments such as dyslexia or stuttering.

The study has three main components:

1. A language background questionnaire
2. A series of short language proficiency tasks
3. Three picture naming tasks

Completeing all tasks will take around 4 hours, divided across three days.
The study is run by PhD-student Malin Mangersnes (malintm@uia.no). Please contact me if you have any queries about the study. My research is supervised by Professor Allison Wetterlin (Allison.wetterlin@uia.no) and Professor Linda Wheeldon (linda.r.wheeldon@uia.no).

WHAT IS THE STUDY ABOUT?

This study is designed to investigate aspects of the use of English as a second language, in particular, in the use of English by speakers that have Norwegian as their first language. The study has three components which will be completed both in Norwegian and English:

1. A questionnaire asking questions about your language background and about how you rate your own level of proficiency in different aspects of the languages that you speak. It should take about 20 minutes to complete (English only).
2. A series of simple language tests assessing vocabulary and reading, as well as some tasks involving repeating nonsense words. These tests will take approximately 1 hour to complete (English and Norwegian).
3. Three picture naming tasks in which you name pictures in your languages as fast and accurately as you can. The total duration is approximately 2 to 2.5 hours (English and Norwegian).

If, after having read the information below, you decide to take part in the study please complete the consent form at the end of this document.

The study will collect and record personal information about you. However, all your data will be pooled with that of other participants for statistical analysis. You will never at any time be mentioned as an individual in relation to this study. Your personal data will be assigned a number code related to your name and stored on a non-networked, password protected PC. Only the laboratory directors and experimenters will have access to the key relating your data number to your name. In addition, we will record the responses you produce during the experiment, this includes key strokes and speech. These data will be also be anonymised and treated as described above.

Page 1 / 3 (ELL-2_PHD_Participant information sheet and consent form_Current_FINAL.docx03-2017 v1)

VOLUNTARY PARTICIPATION AND THE POSSIBLITY TO WITHDRAW CONSENT (OPT-OUT)

Participation in the study is voluntary. If you wish to take part, you will need to sign the declaration of consent on the last page of this document. This will allow us to process your data. You can, at any given time and without reason withdraw your consent. If you decide to withdraw participation in the project, you can ask that your test results and personal data be deleted, unless the data and tests have already been analysed or used in scientific publications.

So long as you can be identified in the collected data you have the right to:
- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data.

If you at a later point, wish to withdraw consent or have questions regarding the project, you can contact the me, Malin Mangersnes. Questions about the study or withdrawing consent can also be directed to the University og Agder's Data protection officer Ina Danielsen ina.danielsen@uia.no or NSD (Norsk senter for forskningsdata AS) by email personvernombudet@nsd.no or telephone 55 58 21 17.

WHAT WILL HAPPEN TO YOUR INFORMATION?

The information that is recorded about you will only be used as described in the purpose of the study.

The results derived from the pooled data will be published. In the interest of being open to the scientific community and others interested in this research we would also like, with your permission, to publish the anonymised data to an open access database. If you agree to this, please sign under "publishing anonymised data to open access database" at the end of this document. The decision you make does not affect your eligibility for this study.
All information will be processed and used without your name or personal identification number, or any other information that is directly identifiable to you.

The principal investigator has the responsibility for the daily operations/running of this research project and that any information about you will be handled in a secure manner. Information about you will be anonymised or deleted when the project is finished. The project period lasts until 08.11.2020, but your personal data may be kept for longer if the project period is extended.

FINANCE

In appreciation for your time and effort, you will receive a voucher for Sørbok for 400 NOK on completion of this study. No payment will be received for partial participation.

APPROVAL

Based on an agreement with The University of Agder, NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

CONSENT FOR PARTICIPATING IN THE RESEARCH PROJECT

I AM WILLING TO PARTICIPATE IN THE RESEARCH PROJECT

TITLE: REPRESENTATION AND PROCESSING OF ENGLISH AS A SECOND LANGUAGE

1) I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
2) I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. I understand that I can withdraw my data at any time during the experiment and after completion of the study until the data is analysed.
3) I understand that data collected during the study will be looked at by researchers from the University of Agder. I give permission for these individuals to have access to my data and to use it for research purposes. I understand that my data will be stored anonymously.
4) I agree to take part in the study.

----

date Participant's Signature

Participant's Name (in BLOCK LETTERS)

PUBLISHING ANONYMISED DATA TO OPEN ACCESS DATABASE

I confirm that anonymised data can be uploaded to an open access database.

----

date Participant's Signature

Appendix B: Ethical considerations

The main ethical considerations for this study relate to the participants and the treatment of the collected data. The following section is based on the Guidelines for Research Ethics in the Social Sciences, Humanities, Law, and Theology (The National Committee for Research Ethics in the Social Sciences and the Humanities, 2016). The guidelines are published by The National Committee for Research Ethics in the Social Sciences and the Humanities (NESH), a part of the Norwegian National Research Ethics Committees.

In line with NESH guidelines the participants were provided with neutral information about the purpose of the project, the methods involved, any possible risks or discomforts, how the data will be stored and used, who will have access to the data, and other important consequences of participation. The participants were informed that they may withdraw from the experiment session or have their data deleted without reason and at any point in time, with the exception of pooled, anonymised data that has already been analysed or published. Written informed consent to participation in the study and use of their data in the described research was obtained at the beginning of the first experiment session.

While no possible risks of participating were identified, the duration of the experiment was quite long (three sessions lasting approximately 4.5 hours combined) which might be uncomfortable for some. Participants had the option of taking a short break between each experimental component and were informed that they can notify the experimenter if they need a break outside these scheduled pauses.

The participants were reimbursed with a gift card. The NESH guidelines state that this could be problematic: "Rewarding or paying participants may also influence the informants' motivation to take part in research projects, and may influence the responses provided by the participants, thus constituting a source of error in the data collected" (The National Committee for Research Ethics in the Social Sciences and the Humanities, 2016, p. 15). It is clear that reimbursement might serve as an incentive for people to participate in the study, and it is used to help recruit the high number of participants which the quantitative experimental approach necessitates. Even though the hourly rate for participation is not very high considering local hourly wages (approximately 100 NOK) this stresses the importance of ensuring informed consent so that participants know what they are agreeing to, aside from the reimbursement.

The second concern raised in the above NESH quotation, reimbursement as a possible source of error, also warrants some consideration. In all studies involving participants, and perhaps especially in studies where participants receive some form of reimbursement, researcher expectations or what participants believe to be the researcher's expectations could bias performance. In the study information letter, it is therefore important to describe the purpose of the study and provide enough information so that informed consent is possible, whilst avoiding explicit mention of expected behaviour/hypotheses or placing value on a certain performance or outcome. This also holds for the experimental sessions. The perhaps most interesting data points collected in my study are immediate and thought to be under little to no conscious control from the participant, and payment is therefore not thought to influence those results. Another means of assuring validity in this study, aside from providing information without biasing the participants, is using both self-reported and behavioural measures of proficiency.

Appendix C: Language questionnaires and supplementary information

Original LEAP-Q (Marian et al., 2007), page 1

Northwestern Bilingualism & Psycholinguistics Research Laboratory
Please cite Marian, Blumenfeld, & Kaushanskaya (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.

Language Experience and Proficiency Questionnaire (LEAP-Q)

Last Name		First Name		Today's Date	
Age		Date of Birth		Male ☐	Female ☐

(1) Please list all the languages you know **in order of dominance**:

1	**2**	**3**	**4**	**5**

(2) Please list all the languages you know **in order of acquisition** (your native language first):

1	**2**	**3**	**4**	**5**

(3) Please list what percentage of the time you are *currently* and *on average* exposed to each language.
(*Your percentages should add up to 100%*):

List language here:					
List percentage here:					

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.
(*Your percentages should add up to 100%*):

List language here					
List percentage here:					

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time.
(*Your percentages should add up to 100%*):

List language here					
List percentage here:					

(6) Please name the cultures with which you identify. On a scale from zero to ten, please rate the extent to which you identify with each culture. (Examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc):

List cultures here					
	(click here for scale	(click here for scale	(click here for scale	(click here for scale	(click here for scale

(7) How many years of formal education do you have? ______ ________________________________
Please check your highest education level (or the approximate US equivalent to a degree obtained in another country):

☐ Less than High School ☐ Some College ☐ Masters
☐ High School ☐ College ☐ Ph.D./M.D./J.D.
☐ Professional Training ☐ Some Graduate School ☐ Other:

(8) Date of immigration to the USA, if applicable ___ ____
If you have ever immigrated to another country, please provide name of country and date of immigration here.
__________________ ____

(9) Have you ever had a vision problem ☐, hearing impairment ☐, language disability ☐, or learning disability ☐ ? (Check all applicable). If yes, please explain (including any corrections):
____________________________________ ____

Original LEAP-Q (Marian et al., 2007), page 2. This page is filled out for each of the languages spoken by the participant.

Language:

This is my (please select from pull-down menu) language.

All questions below refer to your knowledge of .

(1) Age when you…:

began acquiring :	*became fluent* in :	*began reading* in :	*became fluent reading* in :

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where is spoken		
A family where is spoken		
A school and/or working environment where is spoken		

(3) On a scale from zero to ten, please select your *level of* ***proficiency*** in speaking, understanding, and reading from the scroll-down menus:

Speaking	(click here for scale	Understanding spoken language	(click here for scale)	Reading	(click here for scale

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning :

Interacting with friends	(click here for pull-down scale)	Language tapes/self instruction	(click here for pull-down scale
Interacting with family	(click here for pull-down scale)	Watching TV	(click here for pull-down scale
Reading	(click here for pull-down scale)	Listening to the radio	(click here for pull-down scale

(5) Please rate to what extent you are currently exposed to in the following contexts:

Interacting with friends	(click here for pull-down scale)	Listening to radio/music	(click here for pull-down scale
Interacting with family	(click here for pull-down scale)	Reading	(click here for pull-down scale
Watching TV	(click here for pull-down scale)	Language-lab/self-instruction	(click here for pull-down scale

(6) In your perception, how much of a foreign accent do you have in ?

(click here for pull-down scale)

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in :

(click here for pull-down scale)

Current questionnaire

Below is a copy of the handout version of the questionnaire. The handout version was made and distributed after data collection in the lab (using the excel version) was completed. The handout contains added instructions (in italics) for difficult questions that sometimes had to be explained by the experimenter when it was completed in the lab. The four sections are presented on separate sheets in the excel document. Changes from the original LEAP-Q are indicated below each page of the questionnaire.

The questions in both versions are identical apart from question 23. In this question participants are asked to list the amount of time spent in different language environments (country, family, school, workplace). For the excel version we separated the last two questions. First asking how long they had spent in a school/workplace where this language is spoken ALL of the time, and then how long they had spent in a school/workplace where this language is spoken SOME of the time. The motivation was to get a better idea of whether people had been in an environment where English was the only language spoken (for instance an English study program, exchange semester or working abroad) or an environment where both English and Norwegian was used. These questions were difficult to answer (e.g., what qualifies as all of the time?) and the responses were not varied enough for the analysis. In the handout it was therefore changed to a school/workplace where this language is spoken all or most of the time.

Participant number:
Date:

Bilingual questionnaire
Thank you for agreeing to filling out this questionnaire. In the following you should write your response when the question is followed by a line, fill in the tables, and check <u>one box only</u> for each question where the response options are listed behind boxes.
In *italics* you will find extra information to help you answer the questions correctly.
Please ask the person who handed you the questionnaire if you have any questions.

<u>Part 1: Screening Questions</u>

1. What is your age? *(in years)* _____________
2. What is your gender? ☐ Male ☐ Female ☐ Nonbinary
3. Are you a native speaker of Norwegian? ☐ Yes ☐ No
4. Is Norwegian the only language you spoke at home growing up (aside from perhaps English)? ☐ Yes ☐ No
5. Are you a reasonably good speaker of English? ☐ Yes ☐ No
6. Do you have normal vision or vision that is corrected to normal with glasses or contact lenses? ☐ Yes ☐ No
7. Can you confirm that you have no language impairments such as dyslexia, stuttering etc.? ☐ Yes ☐ No
8. Do you have normal hearing or hearing that is corrected to normal?
   ☐ Yes ☐ No
9. Are you left or right handed? ☐ Left ☐ Right
10. What is your country of birth? _____________
11. What is your current country of residence? _____________
12. How many years of education do you have? _____________
    (from primary school onwards. "Folkehøyskole" counts.)
13. What is the highest education level you have?
    ☐ Less than high school ☐ High school ☐ Professional training
    ☐ Current bachelor student ☐ Bachelor's degree ☐ Current master student
    ☐ Master's degree ☐ PhD
    ☐ Other: _____________

    (If other, please specify on the line above.)

1/8

Questions 3-5 and 9 -11: added
All: change in wording and structure

Part 2: Language background

14. Please list all (including languages learned at school) the languages you speak in order of DOMINANCE (up to 5)
(Dominance = higher the more you use a language and the better you are at that language).

	Language
1	
2	
3	
4	
5	

15. Please list all the languages you speak in order of ACQUISITION (up to 5). *(Acquisition = when you learned the language)*

	Language
1	
2	
3	
4	
5	

16. Please list what percentage of the time you are on average exposed to each language (e.g. exposure in terms of talking, listening, and reading, including TV, films and music).
(All your answers should add up to 100%)

	Language	Percentage
1		
2		
3		
4		
5		

2/8

Question 14 and 16: change in wording

17. Please list what percentage of the time you spend speaking each language.
(All your answers should add up to 100%)

	Language	Percentage
1		
2		
3		
4		
5		

18. Please list what percent of time you typically spend reading in each language.
(All your answers should add up to 100%)

	Language	Percentage
1		
2		
3		
4		
5		

19. When choosing a language to speak, with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time.
(All your answers should add up to 100%)

(For this question you should think about what language you prefer for speaking when you do not need to consider the person you are talking to. You may choose one of your languages all of the time (100%), or divide the amount of speaking time across your languages)

	Language	Percentage
1		
2		
3		
4		
5		

3/8

Question 17 and 18: added.

20. What cultures do you identify with (e.g., Norwegian, British, American, etc)? Please list each culture below (up to 5) and use the scale from 0-10 to rate the degree of identification, whereby 0 = no identification, 5 = moderate identification, 10 = complete identification.

    (Here you may list one or more cultures you identify with. "Culture" includes, but is not limited to: attitudes, values, language, art, media, society in general)

	Culture	Identification
1		
2		
3		
4		
5		

21. Do you feel that you were once better in one of your languages and that you have become less fluent? ☐ Yes ☐ No

    If yes, which one? _____________
    And at what age did you become less fluent? _____________

22. In which language do you usually do the following tasks?

	Task	Language
1	Simple maths (count, add)	
2	Dream	
3	Express anger or affection	
4	Talk to yourself	

<u>Part 3: Norwegian and English proficiency</u>

23. Please list the number of years and months you have spent in each language environment.

	Norwegian		English	
	Years	Months	Years	Months
A country where this language is spoken				
A family where this language is spoken				
A school where this language is spoken all or most of the time				
A workplace where this language is spoken all or most of the time				

4/8

Question 20: change in wording

Question 21 and 22: added

Question 23: change in wording

24. Please rate how much the following factors contributed to your learning of each language on a scale of 0-10 whereby 0 = not a contributor, 5 = moderate contributor and 10 = most important contributor.

	Norwegian	English
Interacting with friends / colleagues		
Interacting with family		
Reading (e.g., books, magazines, online material)		
School and education		
Self-instruction (e.g., language learning videos or apps)		
Watching TV / streaming		
Listening to music/media		

25. Please rate to what extent you are currently (e.g. in the last month or so) exposed to each language on a scale of 0-10 whereby 0 = never, 5 = half of the time and 10 = almost always.
(Since 5 = half of the time the total should not be more than 10 for each row. E.g. if you use each language an equal amount when interacting with friends you fill in 5 for Norwegian and 5 for English. NOTE: Less than 10 in total is possible if you spend some time using a language other than Norwegian and English or you do not engage in the listed activity.)

	Norwegian	English
Interacting with friends / colleagues		
Interacting with family		
Reading (e.g., books, magazines, online material)		
Self-instruction (e.g., language learning videos or apps)		
Watching TV / streaming		
Listening to music/media		

26. Please rate your level of proficiency in the following aspects of each language on a scale of 0-10 whereby: 0 = none; 1 = very low; 2 = low; 3 = fair; 4 = slightly less than adequate; 5 = adequate; 6 = slightly more than adequate; 7 = good; 8 = very good; 9 = excellent; 10 = perfect.

	Norwegian	English
Speaking (general fluency)		
Pronunciation (accent)		
Reading		
Writing		
Grammar		
Vocabulary		
Spelling		

5/8

Question 24: change in wording and added school and education response category.
Question 25: change in wording and removed "language lab" response category.
Question 26: change in wording. Divided speaking into general fluency and accent. Removed understanding spoken language and added writing, grammar, vocabulary, and spelling.

27. Please list the AGE (in years) you were when the following occurred for each language.
    (Your best guess or a rough estimation is fine here)

	Norwegian	English
Started hearing this language on a regular basis		
Became fluent in speaking this language		
Started learning to read in this language		
Became fluent in reading this language		

Part 4: Dialect and accent

28. Which dialect of Norwegian do you speak? ____________________

29. How important is speaking your own dialect for you on a scale of 0-10 (whereby 0 = not at all, 5 = moderately important, 10 = extremely important)? ___________

30. To what extent would you say you modify your own dialect when speaking to a person with a different dialect on a scale of 0-10 (whereby 0 = not at all, 5 = moderately, 10 = totally)? ____________

31. Have you lived in an environment where you have been exposed to other dialects than your own for a longer period of time (e.g. moving to a different city in Norway or living with someone who speaks another dialect)?
    (More than one? Write dialects and years separated by commas)

    ☐ Yes ☐ No

    If yes, which dialect(s)? ____
    And for how long (in years)? ___________________

32. In your opinion how strongly regional is your spoken Norwegian on a scale of 0-10 (whereby 0 = not at all, 5 = moderately, 10 = very much)? __________
    (i.e. How easily can someone who knows your dialect identify where you are from

33. What kind of accent do you think your spoken English has (e.g., British / American / other / none in particular)? ____

34. In your view, how much of a Norwegian accent do you have when you speak English on a scale of 0-10? Whereby 0 = none, 1 = almost none, 2 = very light, 3 = light, 4 = some, 5 = moderate, 6 = considerable, 7 = heavy, 8 = very heavy, 9 = extremely heavy, 10 = pervasive. __________

6/8

Question 27 and 34: Change in wording

Question 28-33: added

35. To what extent do you think others identify you as a non-native speaker based on your ACCENT when speaking English on a scale of 0-10 (whereby 0 = never, 5 = half of the time 10 = always)? __________

36. How important is it for you to have a good accent when speaking English on a scale of 0-10 (whereby 0 = not at all, 5 = moderately important, 10 = extremely important)? __________

37. How much effort have you put into improving your accent when speaking English on a scale of 0-10 (whereby 0 = no effort at all, 5 = moderate effort, 10 = constant effort)? __________

38. How would you rate your ability to imitate foreign accents and dialects on a scale of 0-10 (whereby 0 = extremely poor, 5 = moderate, 10 = extremely good)?__________

39. Please rate the degree to which you agree with the following statements on a scale of 0-10 (whereby 0 = very strongly disagree, 10 = very strongly agree)?

Statement	Rating
It is important to me to speak grammatically correct English	
I pay attention to how people pronounce words and sounds	
I want to improve my pronunciation of English	
If it were possible, I would like to pronounce English like a native speaker	
Pronunciation is not important to me because it does not affect how well I can communicate	

40. Are there any sounds in the English language you find difficult to pronounce?
☐ Yes ☐ No
If yes, which one(s)?

____

(Write down the letter representing the sound or a word that contains the sound and underline the relevant portion of the word).

41. Have you noticed any English speech sounds that are difficult for other Norwegians when speaking English? ☐ Yes ☐ No

If yes, which one(s)?

____

(Write down the letter representing the sound or a word that contains the sound and underline the relevant portion of the word).

7/8

Question 35: change in wording

Question 36 – 41: added. Question 39 adapted from Flege et al. (1999)

NOTE: **The next two questions are similar, but take care: the first is asking about accidental language mixing and the second about intentional language mixing**

42. When you are speaking do you ever find yourself <u>accidentally</u> mixing words or sentences from Norwegian and English? ☐ Yes ☐ No

    (a) If yes, how often does English accidentally intrude in your Norwegian on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)? ____________

    (b) And how often does Norwegian accidentally intrude into your English on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)? ____________

43. When you are speaking with a person who also knows both Norwegian and English do you ever find yourself <u>intentionally</u> mixing words or sentences from Norwegian and English? ☐ Yes ☐ No

    (a) If yes, how often does English intentionally intrude in your Norwegian on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)? ____________

    (b) And how often does Norwegian intentionally intrude into your English on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)? ____________

44. Which written form of Norwegian have you predominantly been using?
    ☐ Bokmål ☐ Nynorsk

END OF QUESTIONNAIRE - THANK YOU FOR YOUR TIME!

8/8

Question 42 – 44: added

Table C1
Overview of missing responses from LEAP-Q, means and replacement values

Question	*N*	*M*	Value entered	Participant ID
Q2c_Contrib_Reading_Nor	1	7.32	7	294
Q2g_Contrib_Music_Nor	1	3.47	3	280
Q4h_Prof_Speaking_Eng	1	7.72	8	264
Q5b_FluentSpeaking_Age_Nor	1	4.30	4	260
Q5d_FluentReading_Age_Nor	1	8.05	8	270
Q5f_FluentSpeaking_Age_Eng	2	12.78	13	213, 231
Q5h_FluentReading_Age_Eng	1	12.12	12	270
Q5_Regional_Rating	1	6.97	7	217
Q7_Heavy_NorsktoEng_Accent	1	3.38	3	225
Q8_Accent_NonNative_Obvious	1	5.70	6	232
Q12d_Want_Like_Native_Eng	1	8.68	9	204
Q12e_Pronouncation_NOT_import	1	3.18	3	204

Appendix D: Lists of stimuli for objective language tests

Spelling task materials

English	**Norwegian**
vouchers	takknemlig
sincerely	iøynefallende
weird	anerkjenne
pursue	kakerlakk
caution	misvisende
tomorrow	fellesskap
disseminate	holke
thoroughly	viderekomne
receipt	bokstavelig
obtain	dessverre
feasible	skjenkebevilling
surveillance	usaklig
breathe	sinnssyk
imageability	bygde
conscience	rappellere
miscellaneous	verre
maintenance	blant
vengeance	bibliotek
questionnaire	tillitvekkende
approximately	unntatt

Vocabulary task materials

English

	Target	Correct	FoilA	FoilB	FoilC
Synonyms	**Target**	**Correct**	**FoilA**	**FoilB**	**FoilC**
	caprice	whim	cattle	brute	lounge
	baffle	confuse	hide	warp	bully
	ponderous	unwieldy	useless	supportive	thoughtful
	banter	chatting	whispering	denial	beating
	garish	tasteless	spiky	green	bland
	sequin	bead	stamp	sledge	order
	loquacious	talkative	broad	roomy	marshy
	covet	desire	pad	cradle	cave
	acumen	cleverness	blame	spicy	wealth
	drench	soak	raise	erase	flatten
	abide	endure	inhabit	crave	depart
	vocation	occupation	holiday	pronunciation	vocabulary
	gulch	crevasse	swallow	shed	dislike
	cogitate	ponder	achieve	succeed	enquire
	vexatious	effortful	engaging	horrifying	priceless
	peril	danger	shiny	delight	shelter
	feral	savage	hungry	impartial	ugly
	ludicrous	ridiculous	developed	nasty	certain
	brisk	energetic	disposable	section	stern
	truculent	defiant	delicious	juicy	tardy
Antonyms	**Target**	**Correct**	**FoilA**	**FoilB**	**FoilC**
	concerned	uncaring	scarce	misleading	understanding
	timorous	fearless	forestry	funny	emotive
	disdain	admire	unload	misfortune	huge
	acerbic	sweet	itchy	loud	beautiful
	nonplus	enlighten	subtract	gain	disadvantage
	surfeit	lack	southern	excess	fake
	vicious	gentle	slippery	fierce	disobedient
	saunter	rush	fry	punish	daydream
	slipshod	careful	difficult	clumsy	footwear
	umbrage	delight	dungeon	demanding	appeal
	strenuous	effortless	arduous	smooth	tricky
	divulge	conceal	purchase	disclose	smuggle
	loathe	cherish	rejoice	kindle	undress
	querulous	agreeable	feathered	blatant	squeaky
	forgo	acquire	precede	journey	disappear
	conquer	surrender	demand	retain	release
	hovel	palace	float	cloudy	stairwell
	adversity	advantage	delay	grudge	persevere
	alacrity	slowness	annoyance	fog	ingenuity
	penury	wealth	dispatch	cunning	famine

Norwegian

Synonyms	**Target**	**Correct**	**FoilA**	**FoilB**	**FoilC**
	lektyre	lesestoff	leker	hytte	husdyr
	ufortrøden	uforstyrrelig	uforbederlig	ufokusert	fornøden
	noksagt	dumrian	ferdigstilt	selvdyrker	påstand
	lemfeldig	forsiktig	uberegnelig	langsom	frimodig
	febrilsk	hektisk	illevarslende	tilstrekkelig	varmblodig
	brudulje	slagsmål	ekteskap	floke	etterligning
	fjetre	lamme	røpe	legere	finne
	vankelmodig	ubestemt	nådeløs	mangelfull	hyklersk
	attrå	begjære	fornærme	avslå	trampe
	kryste	klemme	brodere	savne	forfølge
	amper	hissig	skyldig	travel	fyldig
	smektende	lengtende	spinkel	smakfull	buktende
	maroder	utmattet	blodtørstig	spenstig	hevngjerrig
	trettekjær	kranglete	grådig	kresen	svak
	fadese	tabbe	utside	krig	vegring
	mulkt	bot	dystert	sveiv	svalt
	atal	plagsom	sløv	dyktig	hvass
	vansmekte	lide	gnage	avsky	forgifte
	sondre	skille	undersøke	forske	vise
	omkalfatre	endevende	oppfatte	omkomme	omlegge
Antonyms	**Target**	**Correct**	**FoilA**	**FoilB**	**FoilC**
	lapidarisk	pratesyk	usann	kortfattet	fremmed
	distré	oppmerksom	utakknemlig	motsatt	fordelt
	sjofel	hyggelig	annerledes	lumpen	skjærende
	vanvidd	fornuft	ordstrid	viktighet	velklang
	armod	rikdom	avsporing	elendighet	bopel
	overflod	fattigdom	omskifte	flom	vrede
	avertere	skjule	tirre	kunngjøre	forstyrre
	nennsom	voldsom	sparsom	virksom	strevsom
	ødsle	spare	hevde	nære	tvile
	bebreide	berømme	beleire	betvile	betenke
	uaffisert	påvirket	redigert	offentlig	merkelig
	besynderlig	alminnelig	snevert	omfattende	anerkjent
	ublu	rimelig	skjør	freidig	skral
	hovmod	ydmykhet	angst	avskjed	tilregnelighet
	anfektelse	visshet	forhindring	åpenbaring	straff
	petimeter	slask	lekmann	tommestokk	skritt-teller
	avferdige	godta	avslutte	forhindre	testamentere
	bifalle	avvise	tilta	snuble	erobre
	fetere	overse	pine	ernære	flytte
	nidkjær	slurvete	trassig	selvopptatt	streng

Elision task materials

English

Repeat	Repeat and remove segment
say /ˈli.ɔgs/	now say /ˈli.ɔgs/ without the /l/
say /θaʊk/	now say /θaʊk/ without the /θ/
say /ˈzæblət/	now say /ˈzæblət/ without the /z/
say /twɛln/	now say /twɛln/ without the /w/
say /ˈsplɔɪtəl/	now say /ˈsplɔɪtəl/ without the /p/
say /ˈskreɪpʊs/	now say /ˈskreɪpʊs/ without the /r/
say /ˈplaɪ.təf/	now say /ˈplaɪ.təf/ without the /f/
say /jaˈlu:m/	now say /jaˈlu:m/ without the /m/
say /ˈtræs.dʒɔɪb/	now say /ˈtræs.dʒɔɪb/ without the /s/
say /ˈæb.sumpt/	now say /ˈæb.sumpt/ without the /m/
say /klɔ:sp/	now say /klɔ:sp/ without the /s/
say /dʒɪlk/	now say /dʒɪlk/ without the /l/
say /ˈfi:knə/	now say /ˈfi:knə/ without the /n/
say /ˈbi:ltrʌm/	now say /ˈbi:ltrʌm/ without the /t/
say /ˈlæn.spʌŋ/	now say /ˈlæn.spʌŋ/ without the /s/
say /ˈpɪlp.sɔɪ/	now say /ˈpɪlp.sɔɪ/ without the /l/
say /ˈrɛmp.slʌf/	now say /ˈrɛmp.slʌf/ without the /m/
say /ˈwɔɪft.nup/	now say /ˈwɔɪft.nup/ without the /t /

Norwegian

Repeat	Repeat and remove segment
si /çɑmt/	si /çɑmt/ uten /ç/
si /pliʀd/	si /pliʀd/ uten /l/
si /plusk/	si /plusk/ uten /s/
si /smeikt/	si /smeikt/ uten /k/
si /sʉmtglen/	si /sʉmtglen/ uten /m/
si /kaŋtʁɑk/	si /kaŋtʁɑk/ uten /t/
si /tenpʁøyt/	si /tenpʁøyt/ uten /p/
si /skʁaʉden/	si /skʁaʉden/ uten /ʁ/
si /lɑpʁi/	si /lɑpʁi/ uten /l/
si /ʁuplet/	si /ʁuplet/ uten /ʁ/
si /stʁøylem/	si /stʁøylem/ uten /t/
si /kʁeifɑg/	si /kʁeifɑg/ uten /g/
si /ʋe.mi:n/	si /ʋe.mi:n/ uten /n/
si /tʁɑpskait/	si /tʁɑpskait/ uten/p/
si /ɔʁbeskt/	si /ɔʁbeskt/ uten /s/
si /fʉknɑt/	si /fʉknɑt/ uten /n/
si /ʀeltsut/	si /ʀeltsut/ uten /l/
si /kømpkɑl/	si /kømpkɑl/ uten /p/

Serial nonword recognition (SNWR) materials

English (phonemic transcriptions)

List 1

<u>1st sequence</u>	<u>2nd sequence</u>	<u>Trial type</u>
bɑɹtʃ tig nʌp gʊk	bɑɹtʃ tig nʌp gʊk	Same
mə˞d gɑɹp tæm pɪb	mə˞d gɑɹp tæm pɪb	Same
kɪb də˞n pʌtʃ gid	kɪb **pʌtʃ də˞n** gid	Different
pim tɑɹg gæb bɑk tʃɛl	pim **gæb tɑɹg** bɑk tʃɛl	Different
gʊt bɑɹg mə˞n nʌk tɛp	gʊt bɑɹg mə˞n nʌk tɛp	Same
tʃum kə˞p lɔɹk nʌg gɑɹn	tʃum kə˞p lɔɹk **gɑɹn nʌg**	Dummy trial
lʊg dʒɑl dɪdʒ kə˞m mɛb	lʊg dʒɑl **kə˞m dɪdʒ** mɛb	Different
kitʃ dʒɑɹn mɛp tə˞g bɪk	kitʃ dʒɑɹn mɛp tə˞g bɪk	Same
kum tɔɹd mɑdʒ dʒʌp gɪk	kum tɔɹd **dʒʌp mɑdʒ** gɪk	Different
lɑd tʌdʒ dʒɪk nɔɹb gɑɹm	lɑd tʌdʒ dʒɪk nɔɹb gɑɹm	Same
kitʃ dʒɑɹn mɛp tə˞g bɪk	kitʃ **mɛp dʒɑɹn** tə˞g bɪk	Different
tɛk kæm mɪtʃ bɑn də˞p	tɛk kæm mɪtʃ bɑn də˞p	Same
tɔɹm pæg jɛk də˞b kɑl bʌp	tɔɹm pæg jɛk də˞b kɑl bʌp	Same
kɔɹp tib nul dʒɑɹk pɪm gə˞tʃ	kɔɹp tib **dʒɑɹk nul** pɪm gə˞tʃ	Different
tə˞dʒ dʒup lɛk nɔrg tʃɪm pib	tə˞dʒ dʒup lɛk nɔrg tʃɪm pib	Same
kə˞n budʒ tʌd lɪg pæb dɔɹt	kə˞n budʒ tʌd lɪg pæb dɔɹt	Same
kɔɹp tib nul dʒɑɹk pɪm gə˞tʃ	kɔɹp tib nul dʒɑɹk **gə˞tʃ pɪm**	Dummy trial
bɔɹdʒ tʃʌd nig dæk kɛb lɑɹm	bɔɹdʒ **nig tʃʌd** dæk kɛb lɑɹm	Different
tɪdʒ mʌp tʃɛn gə˞b nʊg dit	tɪdʒ mʌp **gə˞b tʃɛn** nʊg dit	Different
pædʒ nɑɹp mɑn tʃʌt gub gɛd	pædʒ nɑɹp mɑn tʃʌt gub gɛd	Same
pɑɹn mæb dɔɹdʒ nɑg tʃim dʒɪt	pɑɹn mæb dɔɹdʒ **tʃim nɑg** dʒɪt	Different
gɛl nə˞g lʌd pɑk mitʃ dub dʒæt	gɛl nə˞g lʌd pɑk **dub mitʃ** dʒæt	Different
dʒɑɹm nɛb gə˞p tʃɔɹg mæl tutʃ lɑn	dʒɑɹm nɛb gə˞p **mæl tʃɔɹg** tutʃ lɑn	Different
tɑb gæn dɑɹtʃ tʃul dʒʌk nɔɹd pɛm	tɑb gæn dɑɹtʃ tʃul dʒʌk nɔɹd pɛm	Same
lim kʌg tʃɔɹdʒ dʒə˞t kɛd dɑɹp gɑk	lim kʌg tʃɔɹdʒ dʒə˞t kɛd dɑɹp gɑk	Same
tuk tʃæd lɪdʒ dʒə˞g dɔp nɑɹt gʌb	tuk tʃæd lɪdʒ **dɔp dʒə˞g** nɑɹt gʌb	Different
dʒɑɹm nɛb gə˞p tʃɔɹg mæl tutʃ lɑn	dʒɑɹm nɛb gə˞p tʃɔɹg mæl tutʃ lɑn	Same
lim kʌg tʃɔɹdʒ dʒə˞t kɛd dɑɹp gɑk	lim kʌg tʃɔɹdʒ dʒə˞t kɛd **gɑk dɑɹp**	Dummy trial
tʃɪg næm pɛb gɑp dʒutʃ lɑɹt tid	tʃɪg **pɛb næm** gɑp dʒutʃ lɑɹt tid	Different
kɑɹk nə˞tʃ mɔɹd tʃæm bul lʌb tip	kɑɹk nə˞tʃ mɔɹd tʃæm bul lʌb tip	Same

English (phonemic transcriptions)

List 2

1st sequence	2nd sequence	Trial type
baɹtʃ tig nʌp gʊk	baɹtʃ tig nʌp gʊk	Same
mɚd gaɹp tæm pɪb	mɚd gaɹp tæm pɪb	Same
kɪb dɚn pʌtʃ gid	kɪb **pʌtʃ dɚn** gid	Different
tɛk kæm mɪtʃ ban dɚp	tɛk **mɪtʃ kæm** ban dɚp	Different
tʃum kɚp lɔɹk nʌg gaɹn	tʃum kɚp lɔɹk nʌg gaɹn	Same
pim taɹg gæb bak tʃɛl	pim taɹg gæb bak tʃɛl	Same
kum tɔɹd madʒ dʒʌp gɪk	kum tɔɹd madʒ **gɪk dʒʌp**	Dummy trial
gʊt baɹg mɚn nʌk tɛp	gʊt baɹg **nʌk mɚn** tɛp	Different
lʊg dʒal dɪdʒ kɚm mɛb	lʊg dʒal dɪdʒ kɚm mɛb	Same
tʃum kɚp lɔɹk nʌg gaɹn	tʃum kɚp **nʌg lɔɹk** gaɹn	Different
kum tɔɹd madʒ dʒʌp gɪk	kum tɔɹd madʒ dʒʌp gɪk	Same
lad tʌdʒ dʒɪk nɔɹb gaɹm	lad **dʒɪk tʌdʒ** nɔɹb gaɹm	Different
paɹn mæb dɔɹdʒ nag tʃim dʒɪt	paɹn mæb dɔɹdʒ nag tʃim dʒɪt	Same
pædʒ naɹp man tʃʌt gub gɛd	pædʒ naɹp man **gub tʃʌt** gɛd	Different
kɚn budʒ tʌd lɪg pæb dɔɹt	kɚn **tʌd budʒ** lɪg pæb dɔɹt	Different
tɪdʒ mʌp tʃɛn gɚb nʊg dit	tɪdʒ mʌp tʃɛn gɚb nʊg dit	Same
tɚdʒ dʒup lɛk nɔrg tʃim pib	tɚdʒ dʒup **nɔrg lɛk** tʃim pib	Different
bɔɹdʒ tʃʌd nig dæk kɛb laɹm	bɔɹdʒ tʃʌd nig dæk kɛb laɹm	Same
kɔɹp tib nul dʒaɹk pɪm gɚtʃ	kɔɹp tib nul dʒaɹk pɪm gɚtʃ	Same
tɔɹm pæg jɛk dɚb kal bʌp	tɔɹm **jɛk pæg** dɚb kal bʌp	Different
kɔɹp tib nul dʒaɹk pɪm gɚtʃ	kɔɹp tib nul dʒaɹk **gɚtʃ pɪm**	Dummy trial
lim kʌg tʃɔɹdʒ dʒɝt kɛd daɹp gak	lim kʌg tʃɔɹdʒ **kɛd dʒɝt** daɹp gak	Different
tʃaɹn nɪg kam dʒil gædʒ lɚb mʌn	tʃaɹn nɪg kam dʒil gædʒ lɚb mʌn	Same
tʃɪg næm pɛb gap dʒutʃ laɹt tid	tʃɪg næm pɛb gap dʒutʃ laɹt tid	Same
kaɹk nɚtʃ mɔɹd tʃæm bul lʌb tip	kaɹk nɚtʃ mɔɹd tʃæm **lʌb bul** tip	Different
tuk tʃæd lɪdʒ dʒɝg dɔp naɹt gʌb	tuk tʃæd lɪdʒ dʒɝg dɔp naɹt gʌb	Same
tab gæn daɹtʃ tʃul dʒʌk nɔɹd pɛm	tab gæn **tʃul daɹtʃ** dʒʌk nɔɹd pɛm	Different
gɛl nɚg lʌd pak mitʃ dub dʒæt	gɛl nɚg lʌd pak mitʃ dub dʒæt	Same
lim kʌg tʃɔɹdʒ dʒɝt kɛd daɹp gak	lim kʌg tʃɔɹdʒ dʒɝt kɛd **gak daɹp**	Dummy trial
tʃaɹn nɪg kam dʒil gædʒ lɚb mʌn	tʃaɹn **kam nɪg** dʒil gædʒ lɚb mʌn	Different

Norwegian (orthographic transcriptions)

List 1

1st sequence	2nd sequence	Trial type
nerp gær leis vur	nerp gær leis vur	Same
tårs føyn darg hirk	tårs darg føyn hirk	Different
kjud bårl sjaum jøt	kjud bårl sjaum jøt	Same
jælk frug klai derr meit	jælk frug klai derr meit	Same
biss krøf raim mell skapp	biss raim krøf mell skapp	Different
vark fost spei kæng borb	vark fost spei kæng borb	Same
pås nauv hek tøgg kripp	pås nauv tøgg hek kripp	Different
grøyd tærv svutt leff nisj	grøyd tærv svutt nisj leff	Dummy trial
veir jerm kjasj dåv rai	veir jerm kjasj dåv rai	Same
grøyd tærv svutt leff nisj	grøyd svutt tærv leff nisj	Different
gof skjoi lib hupp fråm	gof lib skjoi hupp fråm	Different
kjusj sløbb dol malk poid	kjusj sløbb dol malk poid	Same
naup kjov pødd gårn væng laff	naup kjov gårn pødd væng laff	Different
lisj dørn gedd plub kræl voig	lisj dørn gedd plub voig kræl	Dummy trial
møng rell bravv skjai vikk grup	møng rell bravv skjai vikk grup	Same
fasj klåff mong rais floi berr	fasj klåff mong rais floi berr	Same
seit hurm jæbb kjauv nang toff	seit hurm kjauv jæbb nang toff	Different
fimm sain møp tus kjeng vasj	fimm sain møp tus kjeng vasj	Same
larg drøs pau nung hob jedd	larg drøs nung pau hob jedd	Different
skjoll høyk bræd teir ripp jåst	skjoll høyk bræd teir ripp jåst	Same
lisj dørn gedd plub kræl voig	lisj gedd dørn plub kræl voig	Different
kell sæbb disj skjo nøyv barr grøk	kell sæbb disj skjo barr nøyv grøk	Different
råb gaut næd serf baur hork plegg	råb gaut næd serf baur hork plegg	Same
hing laid tøp boll sæmm kån skjalt	hing laid tøp boll sæmm kån skjalt	Same
deiv log vøy kang trøpp måsj kjul	deiv log vøy trøpp kang måsj kjul	Different
hobb lårr diff køsj jøyt pærg rudd	hobb lårr diff køsj jøyt pærg rudd	Same
læg frinn rait kjess poin juff håm	læg frinn rait kjess poin juff håm	Same
nåv skjon gøl vumm misj fæss kjoit	nåv skjon gøl misj vumm fæss kjoit	Different
skjøm jång mes filk tat stæl biv	skjøm jång mes filk tat biv stæl	Dummy trial
skjøm jång mes filk tat stæl biv	skjøm jång filk mes tat stæl biv	Different

Norwegian (orthographic transcriptions)

List 2

1st sequence	2nd sequence	Trial type
nerp gær leis vur	nerp gær leis vur	Same
tårs føyn darg hirk	tårs **darg føyn** hirk	Different
kjud bårl sjaum jøt	kjud bårl sjaum jøt	Same
kjusj sløbb dol malk poid	kjusj **dol sløbb** malk poid	Different
biss krøf raim mell skapp	biss krøf raim mell skapp	Same
pås nauv hek tøgg kripp	pås nauv hek tøgg kripp	Same
veir jerm kjasj dåv rai	veir jerm **dåv kjasj** rai	Different
grøyd tærv svutt leff nisj	grøyd tærv svutt leff nisj	Same
jælk frug klai derr meit	jælk **klai frug** derr meit	Different
veir jerm kjasj dåv rai	veir jerm kjasj **rai dåv**	Dummy trial
vark fost spei kæng borb	vark fost **kæng spei** borb	Different
gof skjoi lib hupp fråm	gof skjoi lib hupp fråm	Same
lisj dørn gedd plub kræl voig	lisj dørn gedd plub kræl voig	Same
skjoll høyk bræd teir ripp jåst	skjoll høyk bræd **ripp teir** jåst	Different
møng rell bravv skjai vikk grup	møng **bravv rell** skjai vikk grup	Different
larg drøs pau nung hob jedd	larg drøs pau nung hob jedd	Same
lisj dørn gedd plub kræl voig	lisj dørn gedd plub **voig kræl**	Dummy trial
seit hurm jæbb kjauv nang toff	seit hurm jæbb kjauv nang toff	Same
naup kjov pødd gårn væng laff	naup kjov pødd gårn væng laff	Same
fasj klåff mong rais floi berr	fasj klåff **rais mong** floi berr	Different
fimm sain møp tus kjeng vasj	fimm sain møp **kjeng tus** vasj	Different
deiv log vøy kang trøpp måsj kjul	deiv log vøy kang trøpp måsj kjul	Same
skjøm jång mes filk tat stæl biv	skjøm jång mes filk tat stæl biv	Same
råb gaut næd serf baur hork plegg	råb **næd gaut** serf baur hork plegg	Different
kell sæbb disj skjo nøyv barr grøk	kell sæbb disj skjo nøyv barr grøk	Same
skjøm jång mes filk tat stæl biv	skjøm jång mes filk tat **biv stæl**	Dummy trial
nåv skjon gøl vumm misj fæss kjoit	nåv skjon gøl vumm misj fæss kjoit	Same
hobb lårr diff køsj jøyt pærg rudd	hobb lårr diff køsj **pærg jøyt** rudd	Different
læg frinn rait kjess poin juff håm	læg frinn **kjess rait** poin juff håm	Different
hing laid tøp boll sæmm kån skjalt	hing **tøp laid** boll sæmm kån skjalt	Different

Appendix E: Factor analysis residuals and factor correlations

Figure E1
Residuals from analysis with 10 unrotated factors

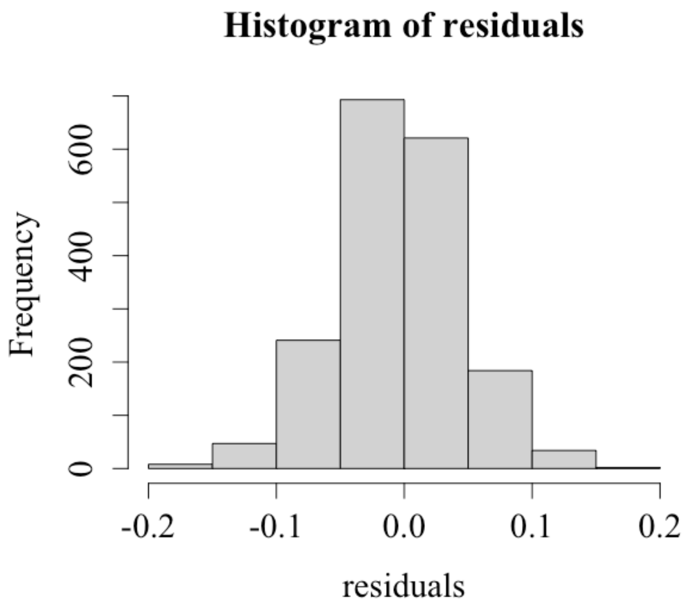


Table E1
Factor correlations from analysis with 10 factors and oblique rotation

	TC6	**TC4**	**TC2**	**TC5**	**TC9**	**TC8**	**TC10**	**TC7**	**TC1**	**TC3**
TC6	1.00	0.21	-0.01	-0.05	0.24	-0.17	0.06	-0.01	-0.07	0.03
TC4	0.21	1.00	0.05	-0.09	0.12	-0.11	0.01	0.04	-0.06	-0.01
TC2	-0.01	0.05	1.00	-0.08	0.05	0.16	0.11	0.02	-0.15	-0.01
TC5	-0.05	-0.09	-0.08	1.00	-0.15	0.05	-0.06	0.05	0.02	0.07
TC9	0.24	0.12	0.05	-0.15	1.00	-0.13	0.07	-0.09	-0.11	-0.03
TC8	-0.17	-0.11	0.16	0.05	-0.13	1.00	0.08	0.07	0.04	0.04
TC10	0.06	0.01	0.11	-0.06	0.07	0.08	1.00	0.07	-0.05	0.11
TC7	-0.01	0.04	0.02	0.05	-0.09	0.07	0.07	1.00	0.05	0.03
TC1	-0.07	-0.06	-0.15	0.02	-0.11	0.04	-0.05	0.05	1.00	0.10
TC3	0.03	-0.01	-0.01	0.07	-0.03	0.04	0.11	0.03	0.10	1.00

Appendix F: Additional factor analyses and results

Four variables were removed from the questionnaire dataset due to low KMO scores. Those were finding L1 dialect important, exposure to L2 through music, interacting with family contributing to L1 learning, and intentionally substituting L1 into the L2. The KMO measure of sampling adequacy in the new dataset was .72 (compared to .70 in the original dataset) and Bartlett's test of sphericity remained significant ($\chi^2(1596) = 5267.52$, $p < .001$). A new correlation matrix was produced for the new dataset. The determinant was slightly improved compared to the original analysis but remained small (6.392181e-15). The results are reported for analyses with orthogonal rotation.

In the factor analysis with the new dataset and 10 factors the root means squared residual was 0.05 and the proportion of residuals exceeding 0.05 was 0.29 (N = 485). The off-diagonal fit was 0.93. In the 9 factor model the root means squared residual was 0.05, the proportion of residuals exceeding 0.05 was 0.30 (N = 484) and the off-diagonal fit was 0.92. The five first factors in the original analysis were very similar to factors found in the new analyses. Extracted factor scores from the original analysis were compared to factor scores from the new analyses, and these were found to correlate highly (see Table F1). The results from the new 10 factor analysis are printed in Table F2 below, followed by the results from the 9 factor analysis in Table F3. Cronbach's alpha and squared multiple correlation (SMC) for each factor are listed below the factor loadings. Factor loadings in bold indicate that this variable was part of the original factor. Variables that were part of the original factor, but not the new factor structure, are marked with an "X".

Table F1
Correlations between original factor scores and factor scores from two new analyses

		New 10 factor analysis		New 9 factor analysis	
Original factor name	**New factor**	r(58)	*p*	r(58)	*p*
General L2 proficiency	RC1	.98	<.001	.98	<.001
L2 accent and interest	RC4	.99	<.001	.99	<.001
L2 exposure and mixing	RC6	.96	<.001	.96	<.001
General L1 proficiency	RC2	.99	<.001	.99	<.001
Language development	RC3	.95	<.001	.93	<.001

Table F2
Results from analysis with new dataset and 10 factors

The top row shows the original name for factors that were similar across analyses. Factor loadings in bold indicate that this variable was part of the original factor. “X” indicates that the variable part of the original, but not the new factor.

	General L2 proficiency	L2 accent and interest	L2 exposure and mixing	General L1 proficiency		Language development				
	RC1	**RC4**	**RC6**	**RC2**	**RC5**	**RC3**	**RC9**	**RC7**	**RC8**	**RC10**
L2 proficiency vocabulary	**0.83**									
L2 proficiency writing	**0.75**									
L2 proficiency pronunciation	**0.74**									
L2 proficiency spelling	**0.73**			**0.40**						
L2 proficiency reading	**0.66**									
Heavy accent	**-0.61**									
Accidental L1 intrusion into L2	**-0.60**									
L2 accent perceived non-native	**-0.43**							-0.33	0.41	
Friend L2 learning	**0.39**									
Reading L2 learning	**0.34**		**0.33**					-0.31		
L2 accent important		**0.76**								
Want to improve L2 pronunciation		**0.73**								
Effort improving L2 accent		**0.72**								
Pay attention to pronunciation		**0.69**								
L2 important speaking grammatically correct	**0.37**	**0.65**								
Want native-like L2 accent		**0.63**								
Pronunciation not important		**-0.54**								
Ability to imitate accents		**0.52**								

	RC1	RC4	RC6	RC2	RC5	RC3	RC9	RC7	RC8	RC10
L2 exposure reading	0.30		**0.68**							
General L2 exposure			**0.65**							
L2 exposure friends			**0.65**							
Time reading L2 (%)	**0.38**		**0.63**							
Time speaking L2 (%)	**0.35**		**0.54**				-0.36			
Intentional L2 mixing into L1			**0.53**							
Accidental L2 intrusion into L1	-0.32		**0.52**							
Choose speaking L2			**0.43**				-0.39			
L1 regional dialect			**-0.42**							
L1 proficiency writing				**0.85**						
L1 proficiency spelling				**0.81**						
L1 proficiency grammar				**0.77**						
L1 proficiency reading				**0.71**						
L1 proficiency vocabulary				**0.67**						
Music L1 learning					0.71					
TV L1 learning					0.66					
L1 exposure TV					0.62		0.35			
L1 exposure music					0.61					
Reading L1 learning					0.57					
School L1 learning					0.44					
L1 exposure reading			**-0.35**		0.43		0.42			
Friend L1 learning			X		0.40					
L2 age fluent speaking						**0.81**				
L2 age fluent reading						**0.73**				
L2 age start hearing language						**0.63**				

	RC1	RC4	RC6	RC2	RC5	RC3	RC9	RC7	RC8	RC10
L2 age start reading						**0.60**				
L1 age fluent speaking						**0.53**				0.35
L1 age fluent reading						**0.49**				0.46
L1 age start reading						**0.45**				0.44
L1 exposure friends			X				0.78			
L1 exposure family							0.71			
Self-instruction L2 learning								0.73		
L2 exposure self-instruction								0.69		
Self-instruction L1 learning					0.33			0.54		
TV L2 learning									0.69	
Music L2 learning									0.60	
L2 Exposure TV			**0.39**				0.39		0.49	
L2 exposure family										0.63
Family L2 learning	X									0.44
	RC1	**RC4**	**RC6**	**RC2**	**RC5**	**RC3**	**RC9**	**RC7**	**RC8**	**RC10**
% variance	0.10	0.07	0.07	0.06	0.06	0.06	0.04	0.04	0.04	0.03
cumulative variance	0.10	0.17	0.23	0.30	0.36	0.42	0.46	0.50	0.54	0.57
Cronbach's α	0.58	0.83	0.71	0.83	0.78	0.75	0.54	0.49	0.5	0.39
SMC	0.9	0.84	0.84	0.84	0.81	0.78	0.76	0.5	0.52	0.48

Table F3

Results from analysis with new dataset and 9 factors

The top row shows the original name for factors that were similar across analyses. Factor loadings in bold indicate that this variable was part of the original factor. "X" indicates that the variable part of the original, but not the new factor.

	General L2 proficiency	L2 accent and interest	General L1 proficiency	L2 exposure and mixing		Language development			
	RC1	**RC4**	**RC2**	**RC6**	**RC5**	**RC3**	**RC9**	**RC7**	**RC8**
L2 proficiency vocabulary	**0.84**								
L2 proficiency writing	**0.77**								
L2 proficiency spelling	**0.73**		**0.40**						
L2 proficiency pronunciation	**0.73**								
L2 proficiency reading	**0.70**								
Heavy accent	**-0.60**								
Accidental L1 intrusion into L2	**-0.57**								
Reading L2 learning	**0.39**			**0.33**					
Friend L2 learning	**0.37**								
L2 accent important		**0.77**							
Effort improving L2 accent		**0.73**							
Want to improve L2 pronunciation		**0.72**							
Pay attention to pronunciation		**0.66**							
L2 important speaking grammatically correct	**0.39**	**0.64**							
Want native-like L2 accent		**0.62**							
Ability to imitate accents		**0.55**							
Pronunciation not important		**-0.52**							

	RC1	RC4	RC2	RC6	RC5	RC3	RC9	RC7	RC8
L1 proficiency writing			**0.85**						
L1 proficiency spelling			**0.81**						
L1 proficiency grammar			**0.78**						
L1 proficiency reading			**0.68**						
L1 proficiency vocabulary			**0.67**						
L2 exposure reading	0.33			**0.67**					
L2 exposure friends				**0.63**					
General L2 exposure				**0.63**					
Time reading L2 (%)	**0.42**			**0.61**					
Intentional L2 mixing into L1				**0.54**					
Accidental L2 intrusion into L1	-0.31			**0.53**					
Time speaking L2 (%)	**0.37**			**0.53**			-0.34	0.34	
Choose speaking L2	**0.30**			**0.43**			-0.37		
L1 regional dialect				**-0.42**					
Music L1 learning					0.71				
L1 exposure TV					0.69				
L1 exposure music					0.65				
TV L1 learning					0.62				
Reading L1 learning					0.50				
L1 exposure reading	-0.31			**-0.35**	0.48		0.39		
Friend L1 learning					0.44				
Self-instruction L1 learning					0.44			0.40	
School L1 learning					0.38				
L2 age fluent speaking						**0.82**			
L2 age fluent reading						**0.73**			

	RC1	RC4	RC2	RC6	RC5	RC3	RC9	RC7	RC8
L2 age start hearing language						**0.65**			
L2 age start reading						**0.60**			
L1 age fluent speaking						**0.49**			
L1 age fluent reading						**0.44**			
Family L2 learning	**0.30**					-0.36			
L1 exposure friends				X			0.78		
L1 exposure family							0.68		
L2 exposure self-instruction								0.67	
Self-instruction L2 learning								0.64	
L2 exposure family								0.44	
L1 age start reading						**0.42**		0.43	
TV L2 learning									0.66
Music L2 learning									0.59
L2 Exposure TV				**0.37**			0.44		0.47
L2 accent perceived non-native	**-0.41**							-0.36	0.45
	RC1	**RC4**	**RC2**	**RC6**	**RC5**	**RC3**	**RC9**	**RC7**	**RC8**
% variance	0.10	0.07	0.07	0.07	0.06	0.06	0.04	0.04	0.04
cumulative variance	0.10	0.17	0.24	0.30	0.37	0.43	0.47	0.51	0.54
Cronbach's α	0.67	0.83	0.83	0.71	0.78	0.73	0.52	0.33	0.5
SMC	0.91	0.84	0.84	0.84	0.81	0.78	0.71	0.58	0.54

Appendix G: Names and item numbers for picture naming stimuli and switching task subset

Noncognates						Cognates					
Picture naming and switching			Picture naming task only			Picture naming and switching			Picture naming task only		
Item number	English	Norwegian	Item number	English	Norwegian	Item number	English	Norwegian	Item number	English	Norwegian
19	turkey	kalkun	12	mermaid	havfrue	1	mouse	mus	17	devil	djevel
26	onion	løk	73	wave	bølge	35	ear	øre	40	clover	kløver
39	pocket	lomme	86	river	elv	75	balloon	ballong	62	wasp	veps
61	newspaper	avis	104	lift	heis	114	fruit	frukt	100	nose	nese
91	mouth	munn	121	maze	labyrint	119	walnut	valnøtt	143	angel	engel
95	suitcase	koffert	164	road	vei	163	coconut	kokosnøtt	238	puzzle	puslespill
122	chair	stol	173	pharmacy	apotek	185	football	fotball	267	knot	knute
139	ghost	spøkelse	184	witch	heks	199	skirt	skjørt	292	compass	kompass
155	bullet	patron	223	zip	glidelås	236	giraffe	sjiraff	300	thread	tråd
239	woman	dame	254	butcher	slakter	245	roots	røtter	317	soldier	soldat
257	cave	hule	272	executioner	bøddel	273	bus	buss	379	zebra	sebra
285	carrot	gulrot	308	dice	terning	319	cow	ku	390	snow	snø
290	cage	bur	313	barber	frisør	376	sculptor	skulptør	416	sock	sokk
362	pool	svømme-basseng	347	money	penger	386	bomb	bombe	420	throne	trone
438	thumb	tommel	382	wallet	lommebok	439	volcano	vulkan	440	statue	statue
466	judge	dommer	448	jar	glass	455	foot	fot	449	bench	benk
487	basket	kurv	453	scissors	saks	480	rose	rose	477	avocado	avokado
526	waitress	servitør	459	mine	gruve	492	nun	nonne	496	circle	sirkel
568	castle	slott	476	apron	forkle	505	book	bok	557	wolf	ulv
569	tooth	tann	478	sink	vask	522	house	hus	593	chocolate	sjokolade
570	bull	okse	563	vulture	gribb	538	jacket	jakke	597	window	vindu
627	sausage	pølse	623	prison	fengsel	546	door	dør	665	panther	panter
707	dog	hund	650	sponge	svamp	610	honey	honning	675	windmill	vindmølle

Note. Item numbers refer to entry in the MultiPic database (Duñabeitia et al., 2018).

Appendix H: Variables controlled in the selection of picture naming and switching stimuli

Table H1

Within-language means of variables controlled in selection of picture naming stimuli, separated by cognate status, and results from paired t-tests

	English				Norwegian		
	Noncognates *M*	Cognates *M*	*p*		Noncognates *M*	Cognates *M*	*p*
BrE word frequency (log10)	4.18	4.16	.849	Word frequency (log10)	3.57	3.54	.873
AmE word frequency (log10)	4.21	4.12	.478				
Number of letters	5.42	5.22	.522	Number of letters	5.44	5.38	.874
Number of syllables	1.70	1.58	.397	Number of syllables	1.86	1.88	.896
Number of phonemes (BrE)	4.74	4.58	.609	Number of phonemes	5.02	4.78	.467
Number of phonemes (AmE)	4.96	4.74	.506				
Visual complexity	2.73	2.62	.268				
Unstressed phonological neighbourhood density (AmE)	13.82	14.87	.753	Phonological neighbourhood density	9.22	9.50	.870
Stressed phonological neighbourhood density (AmE)	13.55	14.23	.832				
Naming agreement (%)	86.07	85.14	.758				

Table H2

Between-language means of variables controlled in selection of picture naming stimuli, separated by cognate status, and results from paired t-tests

	Cognates				**Noncognates**		
	English *M*	Norwegian *M*	*p*		English *M*	Norwegian *M*	*p*
Word frequency (log10) (BrE)	4.22	3.59	<.001	Word frequency (log10) (BrE)	4.25	3.61	<.001
Word frequency (log10) (AmE)	4.19	3.59	<.001	Word frequency (log10) (AmE)	4.28	3.61	<.001
Number of letters	5.33	5.50	.645	Number of letters	5.54	5.56	.954
Number of syllables	1.54	1.85	.042	Number of syllables	1.67	1.83	.281
Number of phonemes (BrE)	4.67	4.88	.545	Number of phonemes (BrE)	4.83	5.13	.369
Number of phonemes (AmE)	4.83	4.88	.905	Number of phonemes (AmE)	5.06	5.13	.853
Unstressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	15.39	9.79	.064	Unstressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	14.30	9.50	.058
Stressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	14.72	9.79	.087	Stressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	14.01	9.50	.072

Table H3
Within-language means of variables controlled in selection of switching task stimuli, separated by cognate status, and results from paired t-tests

	English				Norwegian		
	Noncognates *M*	Cognates *M*	*p*		Noncognates *M*	Cognates *M*	*p*
BrE word frequency (log10)	4.44	4.38	.681	Word frequency (log10)	3.76	3.67	.661
AmE word frequency (log10)	4.39	4.43	.846				
Number of letters	5.00	5.58	.194	Number of letters	5.21	5.42	.727
Number of syllables	1.42	1.63	.248	Number of syllables	1.63	1.83	.291
Number of phonemes (BrE)	4.29	4.83	.267	Number of phonemes	4.46	4.92	.357
Number of phonemes (AmE)	4.46	4.88	.393				
Visual complexity	2.67	2.75	.624				
Unstressed phonological neighbourhood density (AmE)	22.10	11.93	.053	Phonological neighbourhood density	12.08	10.96	.687
Stressed phonological neighbourhood density (AmE)	20.94	11.48	.054				
Naming agreement (%)	92.05	92.87	.824				

Table H4

Between-language means of variables controlled in selection of switching task stimuli, separated by cognate status, and results from paired t-tests

Cognates				Noncognates			
	English *M*	Norwegian *M*	*p*		English *M*	Norwegian *M*	*p*
Word frequency (log10) (BrE)	4.44	3.76	.002	Word frequency (log10) (BrE)	4.38	3.67	<.001
Word frequency (log10) (AmE)	4.39	3.76	.006	Word frequency (log10) (AmE)	4.43	3.67	<.001
Number of letters	5.00	5.21	.687	Number of letters	5.58	5.42	.755
Number of syllables	1.42	1.63	.248	Number of syllables	1.63	1.83	.291
Number of phonemes (BrE)	4.29	4.46	.739	Number of phonemes (BrE)	4.83	4.92	.862
Number of phonemes (AmE)	4.46	4.46	1.000	Number of phonemes (AmE)	4.88	4.92	.930
Unstressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	22.10	12.08	.044	Unstressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	11.93	10.96	.768
Stressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	20.94	12.08	.057	Stressed phonological neighbourhood density (AmE) - Norwegian phonological neighbourhood density	11.48	10.96	.871

Appendix I: Post hoc analysis of onset type in picture naming

Table I1
Model output assessing influence of onset voicing in picture naming task

Model: -1000 *1/RT ~ CognateStatus*NamingLanguage*Session + voice*CognateStatus*NamingLanguage +
(CognateStatus + NamingLanguage + CognateStatus:NamingLanguage|SUBJ_ID) +
(NamingLanguage+Session + NamingLanguage:Session |Item)

	Model summary			Model comparison			B-H
	Estimate	*SE*	*t*	χ^2	df	*p*	sign*
Fixed effects							
(Intercept)	-1.3036	0.0207	-63.067				
CogStatus	-0.0997	0.0242	-4.122	25.94	1	<.001	yes
Language	-0.0311	0.0128	-2.433	5.93	1	.015	yes
Session	-0.0013	0.0108	-0.119	0.06	1	.808	no
voice	-0.0064	0.0149	-0.429	0.18	1	.670	no
CogStatus:Language	-0.0106	0.0172	-0.617	0.15	1	.702	no
CogStatus:Session	-0.0476	0.0102	-4.655	17.22	1	<.001	yes
Language:Session	0.1563	0.0676	2.312	6.87	1	.009	yes
CogStatus:voice	-0.0027	0.0299	-0.090	0.01	1	.916	no
Language:voice	0.0232	0.0180	1.291	1.24	1	.268	no
CogStatus:Language:Session	-0.0291	0.0174	-1.668	2.74	1	.098	no
CogStatus:Language:voice	-0.0334	0.0359	-0.930	0.83	1	.361	no
Random effects	s^2						
Item (Intercept)	0.014						
Subject (Intercept)	0.017						

*Benjamini–Hochberg significance

Table I2

Model output: influence of onset manner of articulation in the picture naming task

Model: -1000 * 1/RT ~ CognateStatus*NamingLanguage*Session + onset*CognateStatus*NamingLanguage + (CognateStatus + NamingLanguage + CognateStatus:NamingLanguage |SUBJ_ID) + (NamingLanguage + Session + NamingLanguage:Session |Item)

	Model summary			Model comparison			B-H
	Estimate	*SE*	*t*	χ^2	df	*p*	sign*
Fixed effects							
(Intercept)	-1.3089	0.0216	-60.602				
CogStatus	-0.1044	0.0273	-3.829	25.09	1	<.001	yes
Language	-0.0332	0.0149	-2.230	8.72	1	.003	yes
Session	-0.0028	0.0108	-0.262	0.09	1	.762	no
onset				3.48	4	.481	no
CogStatus:Language	0.0008	0.0230	0.036	0.01	1	.918	no
CogStatus:Session	-0.0502	0.0103	-4.865	19.13	1	<.001	yes
Language:Session	0.1574	0.0675	2.331	6.94	1	.008	yes
CogStatus:onset				2.70	4	.610	no
Language:onset				3.26	4	.516	no
CogStatus:Language:Session	-0.0275	0.0174	-1.587	2.48	1	.115	no
CogStatus:Language:onset				3.35	4	.500	no
Random effects	s^2						
Item (Intercept)	0.014						
Subject (Intercept)	0.017		*Benjamini–Hochberg significance				

Appendix J: Distribution of residuals in picture naming and language switching

Figure J1
Distribution of residuals from the final picture naming model

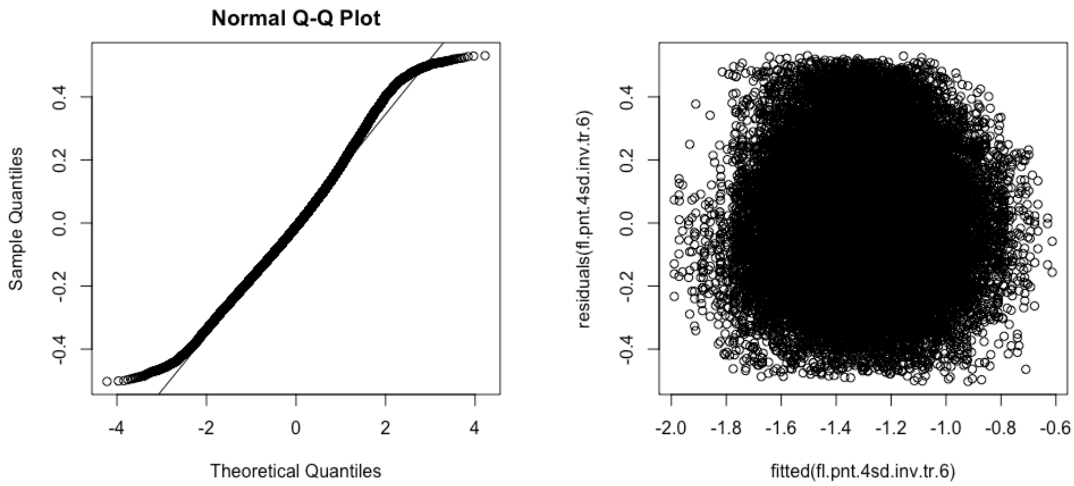
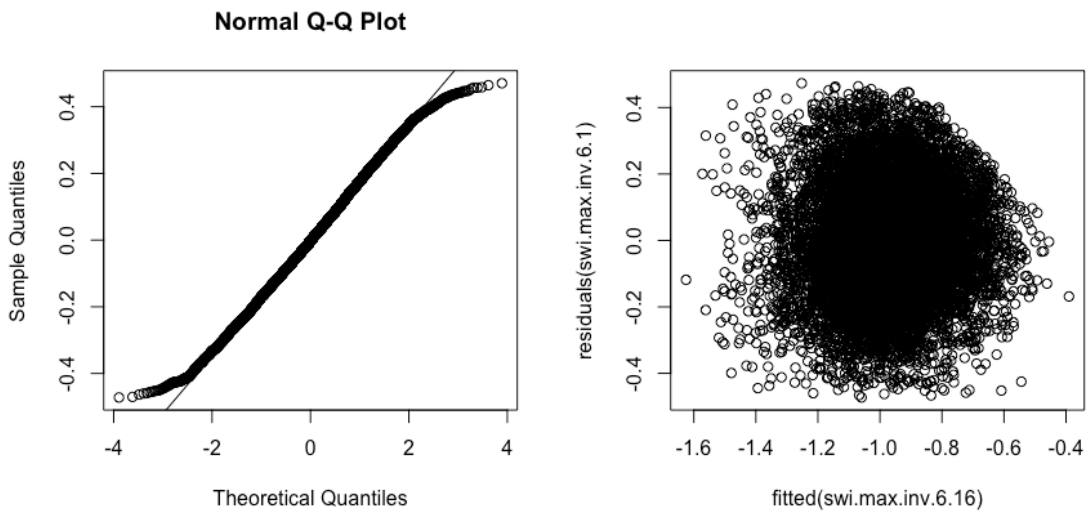


Figure J2
Distribution of residuals from the final language switching model



Appendix K: AIC and BIC for models of reaction times (RT) and error rates in the picture naming and switching task

	Picture naming RT		Picture naming errors		Switching RT		Switching errors	
Model	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
1	-22182	-21829	11080	11289	-5903	-5528	4036	4358
2	-22184	-21839	11078	11278	-5905	-5537	4034	4349
3	-22185	-21848	11076	11268	-5906	-5545	4032	4340
4	-22187	-21859	11075	11258	-5907	-5554	4030	4331
5	-22187	-21868	11075	11241	-5908	-5561	4029	4322
6	-22187	-21876	11074	11230	-5910	-5571	4028	4314
7	-22189	-21887	11072	11220	-5912	-5580	4027	4306
8			11070	11210	-5914	-5589	4027	4298
9			11069	11200	-5916	-5598	4028	4292
10			11069	11191	-5917	-5607	4026	4282
11			11068	11181	-5919	-5616	4024	4274
12					-5921	-5625	4023	4265
13					-5922	-5634	4023	4258
14					-5922	-5640	4021	4248
15					-5924	-5650	4019	4239
16					-5926	-5659	4017	4230
17					-5927	-5668	4015	4221
18					-5929	-5677	4014	4212
19					-5930	-5685	4012	4203
20					-5931	-5693	4013	4196
21							4013	4189
22							4012	4181

Appendix L: Result tables for errors in picture naming and language switching tasks

Table L1
Output of error rate analysis in the picture naming task

	Model summary			Model comparison		
	Estimate	*SE*	*z*	χ^2	df	*p*
Fixed effects						
(Intercept)	-4.1382	0.1376	-30.081			
CogStatus	-0.6496	0.2130	-3.049	8.21	1	.004
Session	-0.5114	0.0798	-6.413	22.11	1	<.001
cAccent	0.0593	0.1007	0.588	0.66	1	.415
Language	-0.2246	0.0823	-2.727	6.08	1	.014
I(cFlanker_diff/1000)	2.7953	1.1901	2.349	4.13	1	.042
CogStatus:Session	-0.2654	0.1137	-2.335	5.24	1	.022
cAccent:Language	-0.2212	0.0896	-2.469	5.61	1	.018
Language:I(cFlanker_diff/1000)	1.6192	0.9328	1.736	2.86	1	.091

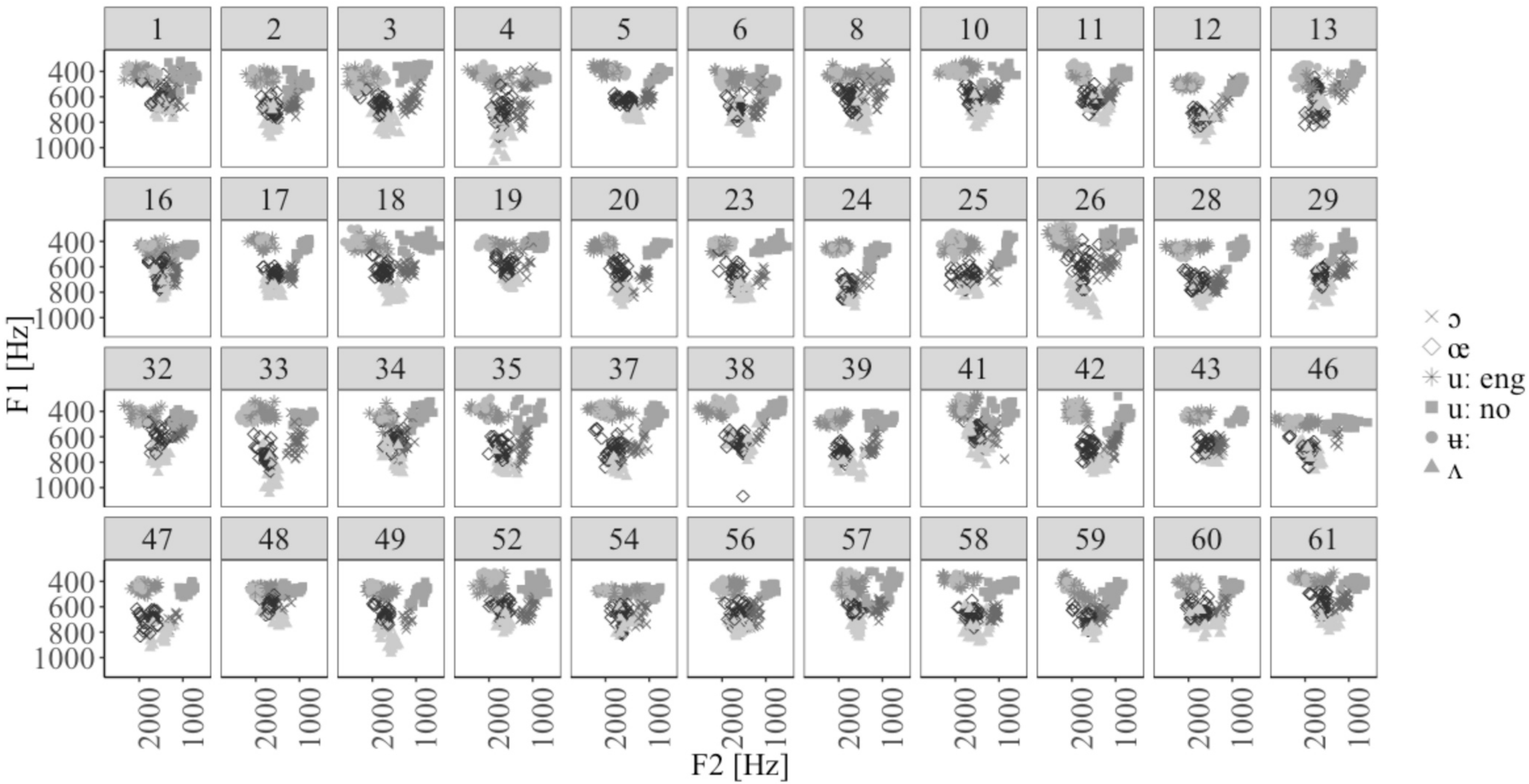
Table L2
Output of error rate analysis in the switching task

	Model summary			Model comparison		
	Estimate	*SE*	*z*	χ^2	df	*p*
Fixed effects						
(Intercept)	-3.5398	0.1632	-21.689			
CogStatus	-0.3312	0.2743	-1.207	1.70	1	.192
Session	-0.2086	0.0894	-2.335	5.98	1	.014
cL2prof	-0.0110	0.1068	-0.187	0.003	1	.955
Language	-0.1098	0.0918	-1.196	1.16	1	.282
TrialType	0.5318	0.0914	5.818	32.62	1	<.001
I(cFlanker_diff/1000)	2.6602	1.3242	2.009	3.84	1	.050
CogStatus:Session	0.3545	0.1788	1.983	3.67	1	.055
CogStatus:cL2prof	0.0397	0.1072	0.371	0.09	1	.766
cL2prof:Language	0.0048	0.0987	0.048	0.02	1	.886
CogStatus:Language	-0.2283	0.1836	-1.244	2.30	1	.129
cL2prof:TrialType	0.1738	0.0988	1.760	3.49	1	.062
CogStatus:TrialType	-0.0843	0.1828	-0.461	0.16	1	.688
Language:TrialType	-0.0708	0.1826	-0.388	0.19	1	.662
CogStatus:L2prof:Language	0.1057	0.1976	0.535	0.0001	1	.994
CogStatus:cL2prof:TrialType	-0.1266	0.1968	-0.643	0.25	1	.614
cL2prof:Language:TrialType	0.0607	0.1969	0.308	0.24	1	.621
CogStatus:Language:TrialType	-0.1599	0.3652	-0.438	0.06	1	.812
CogStatus:cL2prof:Language:TrialType	-1.0719	0.3978	-2.695	6.96	1	.008

Appendix M: Words, vowels, and measurement location for formant analysis

ARPABET	IPA	Language	Cognate	Word	Measurement location
AH	ʌ	English	0	onion	middle
AH	ʌ	English	1	coconut	first half
AH	ʌ	English	0	judge	first half
AH	ʌ	English	1	walnut	first half
AH	ʌ	English	0	thumb	first half
AH	ʌ	English	1	honey	first half
UW	uː	English	0	newspaper	middle
UW	uː	English	1	balloon	middle
UW	uː	English	0	suitcase	middle
UW	uː	English	1	fruit	first half
UW	uː	English	1	roots	first half
UW	uː	English	0	tooth	first half
OX	œ	Norwegian	1	kokosnøtt	first half
OX	œ	Norwegian	1	valnøtt	first half
OX	œ	Norwegian	1	skjørt	first half
OX	œ	Norwegian	0	pølse	first half
OX	œ	Norwegian	0	svømmebasseng	first half
OX	œ	Norwegian	0	kjøkken	first half
UL	uː	Norwegian	0	stol	middle
UL	uː	Norwegian	1	rose	middle
UL	uː	Norwegian	0	gulrot	first half
UL	uː	Norwegian	1	fot	first half
UL	uː	Norwegian	1	bok	first half
UL	uː	Norwegian	0	patron	first half
UX	ʉː	Norwegian	0	bur	first half
UX	ʉː	Norwegian	0	kalkun	first half
UX	ʉː	Norwegian	1	hus	first half
UX	ʉː	Norwegian	1	ku	first half
UX	ʉː	Norwegian	0	hule	first half
UX	ʉː	Norwegian	1	mus	first half
AO	ɔ	Norwegian	1	nonne	middle
AO	ɔ	Norwegian	1	ballong	middle
AO	ɔ	Norwegian	0	dommer	middle
AO	ɔ	Norwegian	0	tommel	middle
AO	ɔ	Norwegian	0	slott	first half
AO	ɔ	Norwegian	1	honning	first half

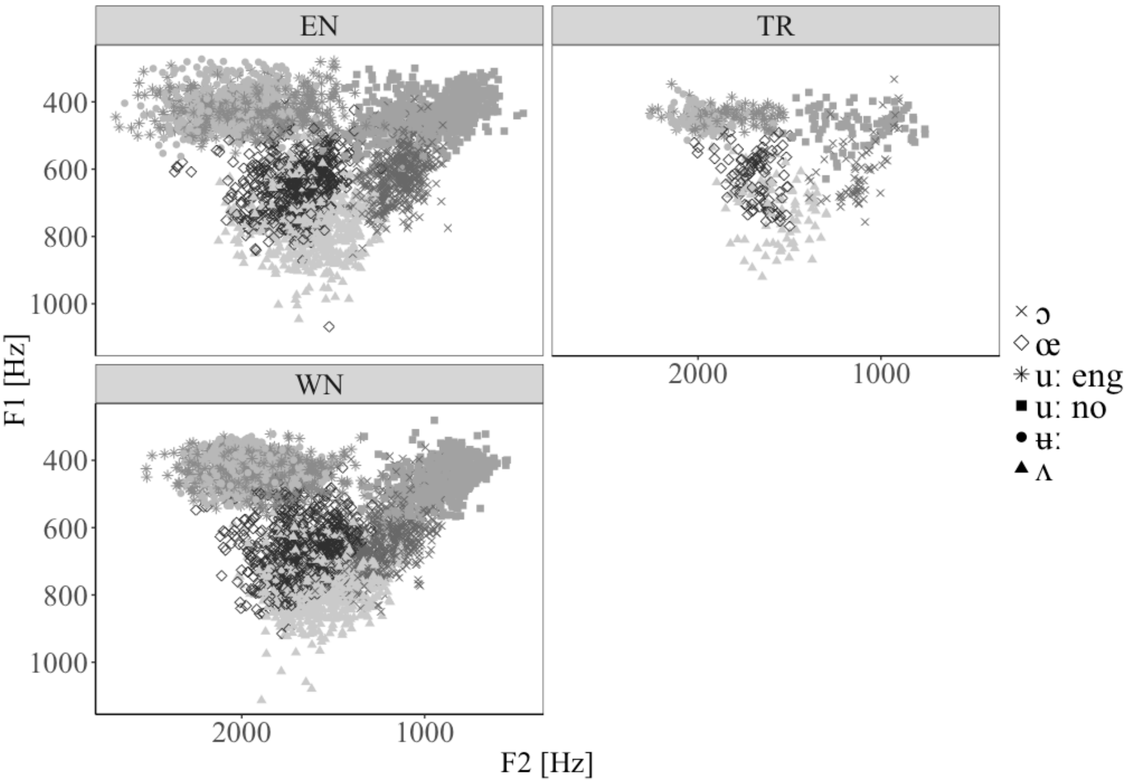
Appendix N: Individual vowel spaces for female speakers



Note. "uː eng" refers to English /uː/ and "uː no" refers to Norwegian /uː.

Appendix O: Vowel spaces divided by dialect group and speaker gender

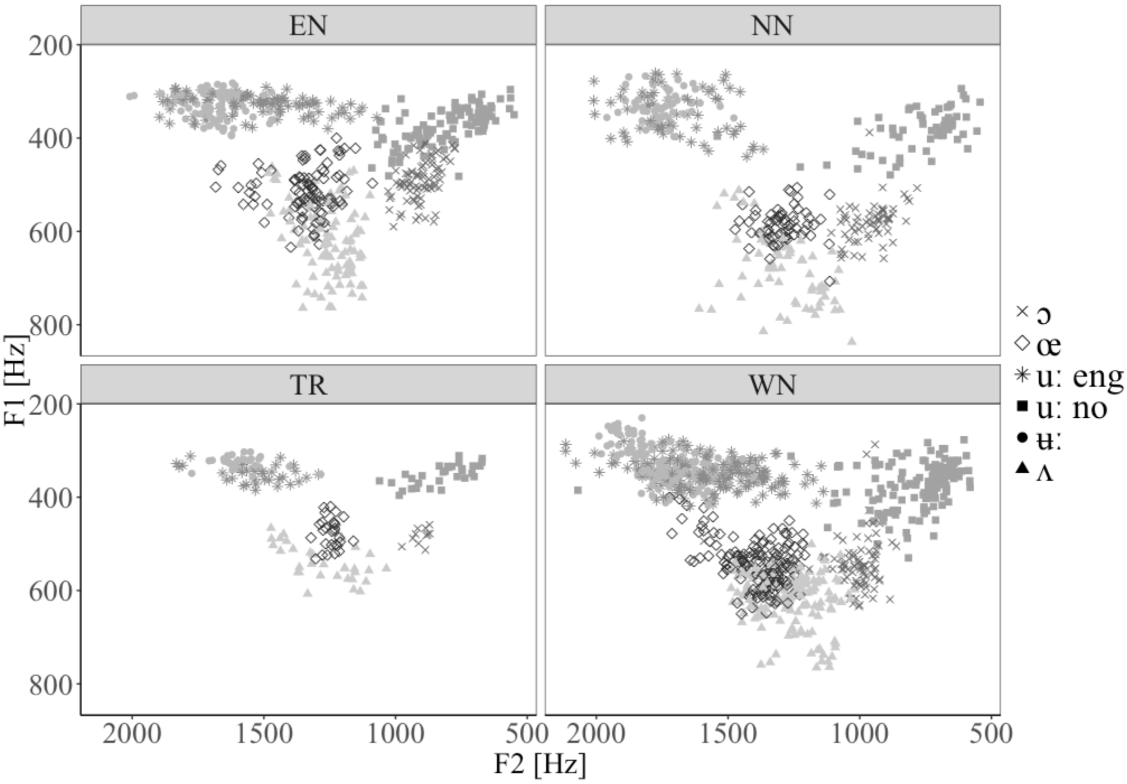
Figure O1
Final data for female speakers by dialect group



Note. Speakers are divided into three regional groups: Eastern Norwegian (EN), from the middle of Norway (TR) and Western Norway (WN). "uː eng" refers to English /uː/ and "uː no" refers to Norwegian /uː/.

Figure O2

Final data for male speakers by dialect group



Note. Speakers are divided into four regional groups: Eastern Norwegian (EN), Northern Norwegian (NN), from the middle of Norway (TR) and Western Norway (WN). "u: eng" refers to English /u:/ and "u: no" refers to Norwegian /u:/.

Appendix P: Results comparing divergence measures to pronunciation proficiency and language test scores

Table P1
Results from correlations with self-rated pronunciation proficiency in L1 and L2 against the four divergence measures

		Pillai score				Bhattacharyya's Affinity			
	Vowel	Pearson		Kendall		Pearson		Kendall	
	contrast	*r*(53)	*P*	r_τ	*p*	*r*(53)	*p*	r_τ	*p*
L1 proficiency	/uː/ - /ʉː/	-.03	.842	.14	.217	-.01	.925	-0.03	.794
	/ʌ/ - /œ/	.12	.402	-.02	.836	-.08	.575	.005	.964
L2 proficiency	/uː/ - /ʉː/	.04	.754	.07	.455	-.15	.271	-.12	.227
	/ʌ/ - /œ/	.10	.485	.11	.289	-.05	.709	-.04	.665

Table P2

Results from multiple linear regression with L1 and L2 test scores against the four divergence measures

		Pillai /uː/ - /ʉː/				Bhattacharyya's Affinity /uː/ - /ʉː/				Pillai /ʌ/ - /œ/				Bhattacharyya's Affinity /ʌ/ - /œ/				
	Test	Est.	*SE*	*t*	*p*	Est.	*SE*	*t*	*p*	Est.	*SE*	*t*	*p*	Est.	*SE*	*t*	*p*	*VIF*
L1	Elision	-0.3844	0.3478	-1.12	.274	0.4187	0.2554	1.64	.107	0.5805	0.3612	1.61	.114	-0.0483	0.2938	-0.17	.870	1.065796
	SNWR	**0.6358**	**0.2448**	**2.60**	**.012**	**-0.6034**	**0.1798**	**-3.36**	**.002**	-0.0046	0.2542	-0.02	.986	0.0105	0.2068	0.05	.960	1.027154
	Spell	0.1284	0.1904	0.67	.503	-0.1838	0.1398	-1.32	.195	0.2173	0.1977	1.10	.277	-0.1089	0.1608	-0.68	.502	1.155103
	Vocab	0.1585	0.1848	0.86	.395	0.0144	0.1357	0.11	.916	0.1748	0.1918	0.91	.366	-0.0531	0.1561	-0.34	.735	1.113172
	Equation	$F(4,50) = 2.37$, $p = .065$, $aR^2 = .092$				$F(4,50) = 3.79$, $p = .009$, $aR^2 = .171$				$F(4,50) = 1.77$, $p = .150$, $aR^2 = .054$				$F(4,50) = 0.23$, $p = .922$, $aR^2 = -.061$				
L2	Elision	-0.3681	0.3664	-1.01	.320	0.3234	0.2842	1.14	.261	0.1668	0.3580	0.47	.643	-0.0501	0.2857	-0.18	.862	1.342212
	SNWR	-0.1503	0.2423	-0.62	.538	0.1304	0.1879	0.69	.491	0.2236	0.2367	0.95	.349	-0.1862	0.1889	-0.97	.329	1.096828
	Spell	0.2381	0.1263	1.89	.065	-0.1361	0.0980	-1.39	.171	**0.2561**	**0.1234**	**2.07**	**.043**	-0.1085	0.0985	-1.10	.276	1.351727
	Vocab	-0.1126	0.1761	-0.64	.525	0.0391	0.1365	0.29	.776	0.0040	0.1720	0.02	.982	-0.0832	0.1373	-0.61	.547	1.283886
	Equation	$F(4,50) = 1.11$, $p = .365$, $aR^2 = .008$				$F(4,50) = 0.86$, $p = .496$, $aR^2 = -.012$				$F(4,50) = 2.24$, $p = .078$, $aR^2 = .084$				$F(4,50) = 1.17$, $p = .337$, $aR^2 = .012$				

Note. SNWR = Serial nonword recognition, Spell = Spelling, Vocab = Vocabulary. Significant and near significant results in bold. Equation = Significance of regression equation and Adjusted R^2 (aR^2). Est. = Estimate. Significant results in bold and marginal result in grey.

Appendix Q: Distribution of residuals and BIC and AIC table for picture naming and language switching with divergence predictors

Figure Q1

Residual plot for the final picture naming task model with the Pillai /uː/-/ʉː/ predictor

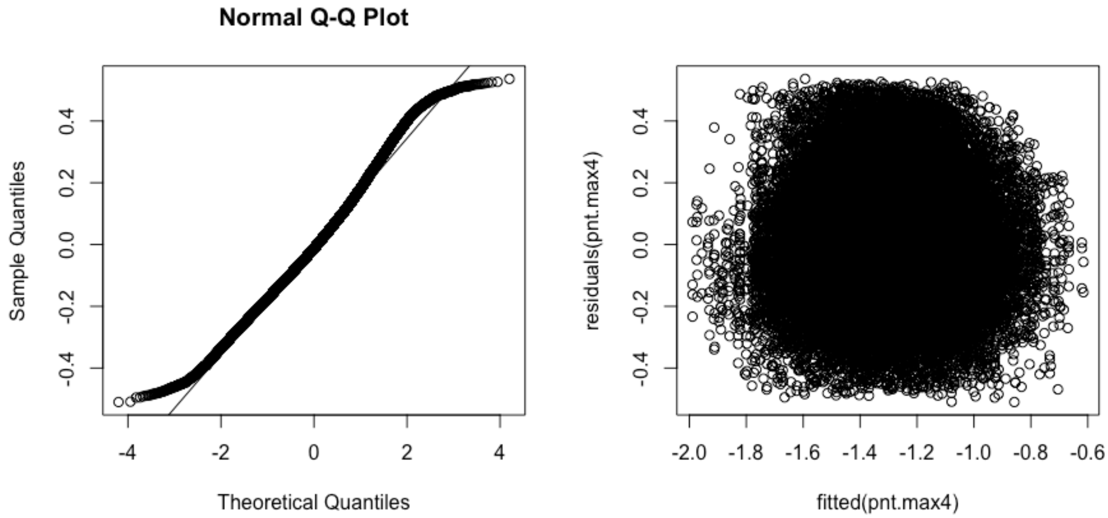


Figure Q2

Residual plot for the final picture naming task model with BA /ʌ/-/œ/ predictor

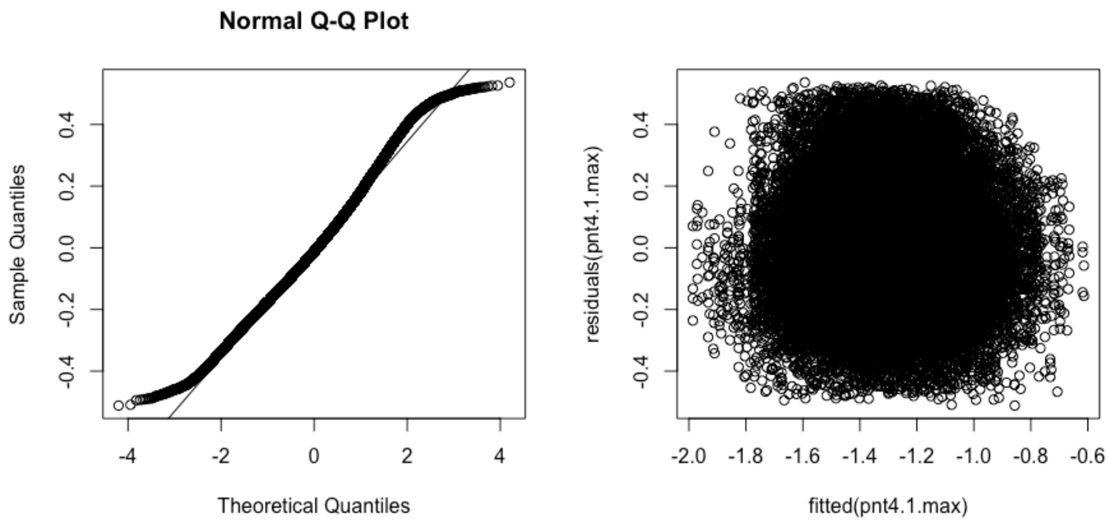


Figure Q3

Residual plot for the final language switching task model with Pillai /uː/-/ʉː/ predictor

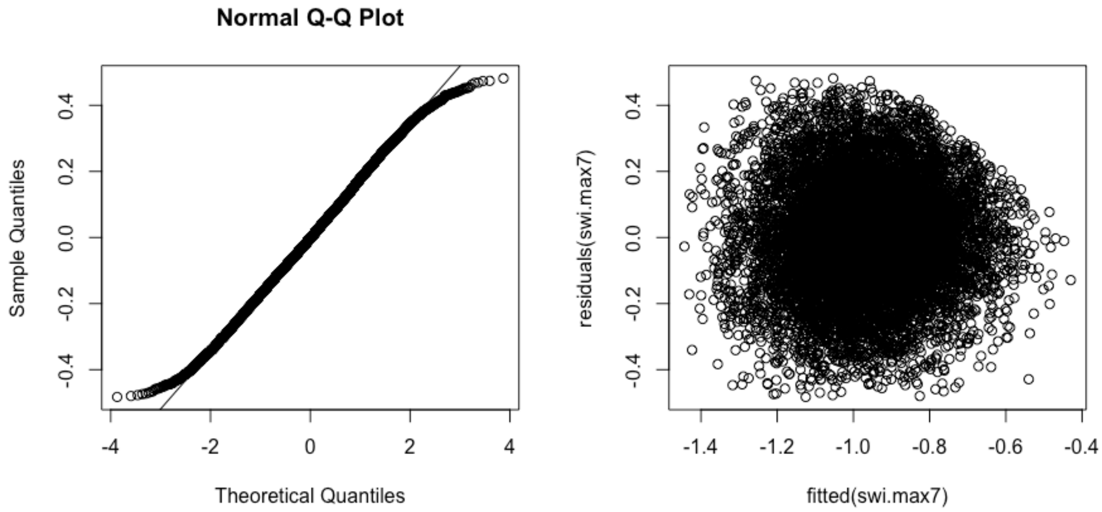


Table Q1

BIC and AIC for model reductions of picture naming and language switching models with divergence measure predictors

	Picture naming				Language switching	
	Pillai score /uː/-/ʉː/		Bhattacharyya's Affinity /ʌ/-/œ/		Pillai score /uː/-/ʉː/	
Model	AIC	BIC	AIC	BIC	AIC	BIC
1	-20367.30	-20085.46	-20364.03	-20082.20	-5564.25	-5307.64
2	-20367.57	-20094.27	-20364.2	-20090.9	-5565.94	-5316.46
3	-20367.37	-20102.62			-5567.64	-5325.28
4	-20367.94	-20111.73			-5569.08	-5333.86
5	-20369.90	-20122.22			-5570.92	-5342.82
6					-5572.30	-5351.33
7					-5572.61	-5358.76
8					-5573.61	-5366.89