Bridging Linguistic Gaps:
The Effects of Linguistic Distance on the Adult Learnability of Dutch as an Additional Language

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: Extract of the Voynich manuscript. Courtesy of the Beinecke Rare Book and Manuscript Library, Yale University.

ISBN: 978-94-6093-165-9
NUR 616

Bridging Linguistic Gaps: The Effects of Linguistic Distance on the Adult Learnability of Dutch as an Additional Language

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 26 maart 2015
om 12.30 uur precies

door

Job Johannes Schepens
geboren op 16 februari 1986
te Nijmegen

Promotor
Prof. dr. R.W.M.N. van Hout

Copromotor
Dr. F.W.P. van der Slik

Manuscriptcommissie
Prof. dr. P.C. Muysken
Prof. dr. M. Dunn (Uppsala Universitet, Zweden)
Prof. dr. B. Szmrecsanyi (Katholieke Universiteit Leuven, België)

Contents

Chapter 1 – Introduction
Chapter 2 – Lexical Distance
Chapter 3 – Morphological Complexity
Chapter 4 – Phonological Distance
Chapter 5 – L1 and L2 Variation
Chapter 6 – The L1 and L2 Distance Effects
Chapter 7 – Discussion and Conclusions

Outline

	Contents	V
	Outline	VII
	Acknowledgements	XIII
1	**Introduction**	**1**
	Adult L2 Learning Difficulty	6
	Linguistic Distance	10
	Lexical Distance (Chapter 2)	11
	Morphological Complexity (Chapter 3)	14
	Phonological Distance (Chapter 4)	16
	Variation in Adult L3 Learnability across L1s and L2s	18
	L3 Learning (Chapter 5)	18
	The L1 and L2 Distance Effects (Chapter 6)	20
	Summary of Objectives	21
	A Description of STEX	22
	Test Design	23
	Variables	24
	Selection Steps	25
2	**The Effect of Linguistic Distance across Indo-European Mother Tongues on Learning Dutch as a Second Language**	**27**
	Introduction	29
	Background	30
	Approaches to Measuring Linguistic Distance	30
	Second Language Learning Effects of Linguistic Distances	33
	Present Study	35
	Methods	37
	Sample	37
	Dependent Variable	38
	Contextual Characteristics	38
	Individual Characteristics	41
	Description of the Sample	42

Analyses 42
Results 45
Estimated Models 46
Fixed Predicted Scores 52
Model Parameter Comparison 56
Discussion and Conclusion 57
3 **Learning Complex Features: A Morphological Account of L2 Learnability** **61**
Introduction 63
Background 65
Methods and Data 70
Proficiency Scores of L2 Dutch 70
Morphological Feature Values 75
Data Analysis 76
Results 78
Feature Patterns 78
Combining Feature Patterns 83
Family Bias 86
Lexical Distance 88
Discussion and Conclusion 90
4 **Learning New Sounds: A Phonological Account of L2 Learnability** **95**
Introduction 97
Present Study 99
Methods 101
Data 101
Sound Inventories 102
Distinctive Features 104
Analysis 105
Study 1 105
Phonological Distance Based on Sound Inventories 106
Measuring Distance between Sound Inventories 106
Example 107
Results 108

Study 2 110
Phonological Distance as based on Distinctive Features 110
Measuring Distance between New Sounds and L1 Sounds 111
Example 113
Results 116
Study 3 118
Comparing Linguistic Distances 118
Discussion 120
Learning Distant Sounds 120
Distance-based Models of L2 Learnability 121
Phonological, Lexical, and Morphological Distance 123
Conclusions 124
5 The L2 impact on Acquiring Dutch as an L3: the L2 Distance Effect 125
Introduction 127
Background 129
CCREMs with Interrelated Random Effects 129
Interrelated L1 and L2 Effects 130
Methods 131
Results 136
Model Comparison 136
Control Variables 140
The L2 Distance Effect 141
Discussion and Conclusion 149
6 L1 and L2 Distance Effects in Learning L3 Dutch 151
Introduction 153
Background 157
L3 Learnability 157
Linguistic Distance 159
Large Scale Studies of Speaking Proficiency 162
Methods 163
STEX 163

        Sample 164
        Task 165
        Measures 165
        Analysis 170
    Results 173
        L1 Distance Effects in Monolingual and Multilingual Learners 173
        Adding L2 Effects for Multilingual Speakers 177
        A Non-Additive L2 Effect: Typological Primacy. 179
        The L1 x L2 Interaction Effect 180
        Effects of L2 Distance or Multilingualism? 182
    Discussion 185
        Consistency with Existing Theories of L3 Learnability 186
        Evaluation of L3 Performance Measures 187
        L1 x L2 Mixing 189
        Linguistic Distance 189
    Conclusion 190
7 Discussion and Conclusions 193
    Summary of Results 194
        The Effect of Linguistic Distance 194
        The Effects of L1 and L2 Distance 196
        General Summary of Results 197
    Approach 197
        Language Testing 197
        Statistical Modeling 201
        Typological Data 204
    Importance 205
        L2 Learnability 205
        L3 Learnability 207
        Learnability Constraints 208
        Third Factors in L2 Learnability 212
    Future Research 215
    Conclusions 218

8 **Appendix A** **221**
9 **Appendix B** **225**
10 **References** **231**
11 **Samenvatting in het Nederlands** **265**
12 **CV** **273**

Acknowledgements

Many people contributed substantially to this dissertation. First, I would like to give my supervisors Frans van der Slik and Roeland van Hout deep credit because I received excellent supervision. Most importantly, we had a great time doing research together. Frans, your daily advice helped me many times to stay focused and to not get stuck on dead ends. I felt comfortable to carry out risky experiments under your supervision. Roeland, you managed to always be one step ahead of me in my own web of ideas and you challenged, helped, and advised me throughout my project. I am also grateful to the members of my doctoral thesis committee. In the final phase of my project, I started collaborating with Florian Jaeger and the people in his lab. This experience also had a positive impact on me and my PhD project and I learned a lot from it. Thanks also to Lou Boves who critically commented on my grant proposal. I decided to start this project based on advice by Ton Dijkstra. Ton, many thanks for your support.

This project benefited from contributions by many people from many PI groups. In Nijmegen: Odette Scharenborg, Antal van den Bosch, Louis ten Bosch. In Rochester: Bozena Pajak, Chigusa Kuramada, Anne-Pier Salverda, Steve Piantadosi, Scott Grimm. In Leipzig: Søren Wichmann, Taraka Rama, Martin Haspelmath, Bernhard Comrie. And in other places: Steven Moran, Michael Cysouw, Shravan Vasisth, Harald Clahsen, Volker Gast, and many others, not to mention organizers of conferences and workshops. A big thanks in particular to the young researchers who inspired this work. In Nijmegen & Saalbach: Steve & Sophie, Fransisco, Jeremy & Kat (& Justin), Ellen & Pascal (& Nora), Saskia & Binyam, Dan & Katie; In Nijmegen: Vanja & Eric, Remy, Claire, Polina, Christel, Maaske, Loes, Richard, Inge, Florian, Ali, Rahim, Sarah, Mira, Harald, Linda, Joshua, Pablo, Louise, Maarten & Anna, Christina & Chris, April Wang, and the students in the statistics courses; In the Nijmegen soccer and basketball teams: Bart, Yang, Robin, Joost, Michele; In Rochester: Amanda Pogue, Esteban Buz, Dave Kleinschmidt, Masha Fedzechkina; and in Leipzig: Linda Gerlach.

The IMPRS, GSH, and LOT research and summer schools have also helped me, much due to research support from Dirkje van der Aa, Tanja Doeller, Els den Os, and Laura Pander. I also made thankful use of a great research infrastructure consisting of the Centre for Language Studies, the Max Planck Institute for Psycholinguistics, the Donders Institute for Cognition (Nijmegen), the Department for Brain and Cognitive Sciences at the University of Rochester, the Department of Linguistics at the MPI for Evolutionary Anthropology, and the Staatsbibliothek zu Berlin. This project made grateful use of data from the state examination of Dutch as an L2. Thanks to Simon Verhallen of the Bureau van het College voor het Toetsen van Examens (CvTE). For use of typological data, I thank Steven Moran, author of PHOIBLE, Matthew Dryer and Martin Haspelmath, authors of WALS, Søren Wichmann, author of ASJP, and Quentin Atkinson, coauthor of Gray & Atkinson (2003). The Centre of Language Studies of the Radboud University primarily supported me, but I received additional funding from the Fulbright program and the Erasmus program. For language help, I thank Theo Bongaerts, Kate Bellamy, Sanne van Vuren, and Martin, Malte, and Thordis. For statistical help, I thank the online course of the University of Bristol, Centre for Multilevel Modelling, and courses by Dan Dediu, Roger Mundry, and Roger Levy. For data collection, I thank a student assistant of the University of Rochester. They are not responsible for any errors in this work.

On a personal note, I am grateful to Mariet, Jan, and Jonas for supporting me throughout. Lastly, I thank Bonnie for her loving support.

Berlin, January 29, 2015 Job Schepens

Chapter 1

Introduction

When we start learning one of the many different languages of the world, we may quickly experience differences in elegance, saliency, complexity, etc. In accordance with this diversity, children seem to have a capacity to use and learn language that has not evolved as a monolithic competence but as a multicomponent enterprise (Darwin, 1871). Child language learners seem to acquire their first as well as additional languages in a generally successful way that is faithful to the language use of the people with whom they communicate. Adult learners of a second or additional language (L2 / L3 / Ln), however, often fail to acquire all of the novelties and peculiarities in the new set of idioms, the different melody, and the grammar. They struggle to produce language that fits the boundaries and constraints of the standards and norms of the L2 native speakers. This research project studied what factors determine the learning difficulties that adult L2 learners have.

We investigated the role of the L1 in learning an additional language by adult learners from the perspective of the typological diversity across the L1s of the learners. Specifically, we investigated what similarities and dissimilarities correlate with differences in adult learnability of an additional language across a wide range of L1s and, at a second stage, across other previously acquired languages. We define L2 learnability as the difficulty to learn an L2 conditional on the L1. In general, the learnability of an additional language is the difficulty to learn that language given all previously learned languages. Thus, the learnability of the words, sounds, and structures of an additional language varies depending on the words, sounds, and structures from previously acquired languages. Similarities and dissimilarities between L1s and languages acquired later allow learners to make inferences about the new, additional language in order to bridge linguistic gaps. For example, given that *tomaat* is a Dutch word, an L2 learner of Dutch first needs to learn the word *tomaat* before he / she can adequately use it. The L2 learnability of *tomaat* may be higher, i.e. it is less difficult to learn the meaning of *tomaat*, when the word already exists in the L1 or is similar to a word in the L1.

The studies presented in this thesis focus on the learnability of Dutch as an additional language, which can be the L2, L3, or Ln, in adult learners who took part in the state exam “Dutch as a Second Language” between 1995 and 2010. The language background of the learners is highly diverse and comprises 74 languages.

Being able to speak the new language of a host society is a crucial part of the human capital of immigrants as it offers them the possibility to communicate and to integrate (Espenshade & Fu, 1997). Governments often require immigrants to learn the dominant language of the country by requiring them to take part in state examinations. Worldwide, learners spend a lot of time and effort in learning new languages. Being proficient in a new or second language is of immediate use in daily life and business, not only for immigrants, but also for everyone else: About 50% of the EU’s population is multilingual (EC, 2012).[1] The L2 (e-)learning industry is one of the fastest growing markets today and has an estimated $50B market size (Baaij, 2012; Rehm & Uszkoreit, 2013), reflecting the societal importance of an L2 and the money that learners want to invest, including these languages which have an international, lingua franca position, English in particular.

Our claim is that the amount of time and effort invested crucially relates to the L1-L2 language (dis)similarities involved. Underestimating this factor may potentially lead to misguided beliefs about language learning capacities of specific groups of adult language learners. L1-dependent courses are not common in the language learning industry, although the need for L1-dependent instructional language learning text books has long been recognized (Lado, 1957) and is still being argued for today (Cook, 2013, p. 180; Macaro, 2006).

[1] See also: Guide for the Development of Language Education Policies in Europe:
http://www.coe.int/t/dg4/linguistic/Source/Guide_Main_Beacco2007_EN.doc

Learners can make significant progress in specialized courses that take L1-L2 (dis)similarities into account. For example, the language center of the Radboud University Nijmegen offers an intense four-week language course for German native speakers with an absolute minimal knowledge of Dutch[2] that prepares them for the state exam "Dutch as an L2" at the B2 level of the Common European Framework of Reference (Council of Europe, 2001). Although such L1-specific courses are available, we know only little about implications of L1-L2 language similarities for language learning in the classroom. Understanding what features of the language background of the language learner matter for adult L2 learning difficulty might have positive consequences for language learners, the language learning industry, policy makers, and subsequently the economic mobility of immigrants. Currently, immigrants may favor migration to countries with linguistically similar languages due to the required language learning investment (Adsera & Pytlikova, 2012).

Applied linguistics has long recognized effects of linguistic differences on second language learning. Contrastive linguistics in the 1960s aimed to explain linguistic interferences that occur in bilingual situations by precisely localizing the linguistic differences between language varieties (Haugen, 1969; Weinreich, 1963). Research on language interference continued with efforts to describe the relevant linguistic differences between languages for language teaching purposes (Lado, 1957). In contrast to interference effects between the L1 and the L2, it is possible to define transfer of L1 *substrate* to an L2 as a positive, facilitating process that occurs when L1-L2 similarities lead to correct language use (Odlin, 1989; Thomason & Kaufman, 1988). At the same time, however, universalists started to emphasize the prevalence of errors of L2 learners that recur across different L1s and their correlation with language universals (Comrie, 1984; Selinker, 1972). The emergence of the concept of universal grammar, supposedly

[2] See: "Niederländisch für deutschsprachige", http://www.ru.nl/studiereninnimwegen/einschreibung/sprachkurs-ru-nt2/

(partly) accessible to adult learners as well, marginalized the study of L1 influences for decades (Foley & Flynn, 2013).

Again, the language learning debate increasingly focuses on what is transferable between language pairs. Our understanding of the struggle of adult language learners may benefit from understanding what linguistic differences make language learning hard, and how this interacts with e.g. developmental factors (Birdsong, 2006, 2014; Bongaerts, 1999; Johnson & Newport, 1989; Lenneberg, 1967). After a certain critical age, L1-L2 linguistic differences pose a major problem for adult learners (Kellerman & Sharwood Smith, 1986; Odlin, 1989). Less linguistic differences seem to define a condition in which it is less difficult to overcome L2 learning problems, besides other individual and contextual factors such as age, the type and duration of exposure, gender, and education (Piske, MacKay, & Flege, 2001).

For understanding the effects of linguistic differences, a distance measure may prove to be supportive, and perhaps may even be crucial. A distance measure can capture differences between many features simultaneously. Multilingualism research has not yet incorporated and applied measures of linguistic distance for investigating and spelling out effects of linguistic differences on additional language learnability. Fairly recently, large-scale databases from linguistic typology became available as resources to figure out and compute distances between widely different languages in the domains of lexicon, grammar/morphology, and phonology.

What measures of linguistic distance can explain differences in L2 learnability? This thesis approaches this question by investigating variation in Dutch speaking proficiency across more than 50,000 adult learners of Dutch who came to the Netherlands for study or work. We utilize a database of their language-testing scores on the state exam Dutch as an L2, which we will refer to as STEX from here on. This set of speaking proficiency scores enables evaluation of the relative impact of linguistic distance of earlier acquired languages on the learnability of an additional language. We aim to develop measures that disclose the impact of linguistic distance. In addition, we would like to study the relation in the other way around as well. We hope that differences in L2 learnability can shed light on compelling questions about linguistic distance.

We will compare and evaluate several ways to determine linguistic distance that can explain variation in L2 learnability, in different linguistic domains, i.e. lexical, morphological, and phonological. In addition, we test whether adult learners use knowledge about a previously learned L2 for making inferences about the new, additional L3. Our studies have become possible due to the advent of advanced mixed effects regression modeling, which allows for the decomposition of variance across different components at the same time (i.e. L1, L2, and country of birth). In addition, our studies depend on the availability of STEX data and on the availability of databases from linguistic typology like WALS (Dryer & Haspelmath, 2011) for morphology and PHOIBLE (Moran, McCloy, & Wright, 2014) for phonology. We show that L2 learning research can profit from including more predictors and developing a number of formalized distance-based models. We envisage that large-scale databases will increasingly play a pivotal role in studying multilingualism, language diversity, and language acquisition, both on the national and the international level.

The remainder of this introduction is organized in three main sections. In the first section, we discuss how L2 learnability varies across L1s and how language-testing data holds the promise of comparing L2 learnability across more L1s than has been possible before. The subsequent section introduces what distance measures may explain variation in L2 learnability across L1s by outlining the goals of Chapter 2, 3, and 4. The third section introduces Chapter 5 and 6, which test whether distance between an already acquired L2 and a target L3 affects L3 learnability. We end the introduction with a summary of our objectives and a methodological background section. Following this introductory chapter, we include five chapters that present original empirical work, and a final chapter with conclusions and discussion. Chapters 2 to 6 contain separate investigations into sub-questions about the relation between language distance and second language learning. Chapter 7 gives a summary and conclusion, and discusses theoretical implications and future research.

Adult L2 Learning Difficulty

A host of factors determines learning difficulty of an L2. For example, at a young age, learners pick up L2 features more easily than later in life. We discuss the most relevant factors before zooming in on the role of language background and L1-L2 differences. A traditional factor is age of learning onset. A long standing body of literature is available with discussions of the linearity or non-linearity of age onset effects on language learning success (Birdsong, 2006, 2014; Bongaerts, Van Summeren, Planken, & Schils, 1997; R. M. DeKeyser, Alfi-Shabtay, & Ravid, 2010; Johnson & Newport, 1989; Scovel, 2000; Vanhove, 2013). Interestingly, observational data provide no evidence for nonlinear effects of age (Bleakley & Chin, 2010; Chiswick & Miller, 2008; Hakuta, Bialystok, & Wiley, 2003).

Other factors at the level of the individual learner affecting the cognitive setting of the learner are working memory and language learning aptitude (P. Robinson, 1996, 2005). A higher aptitude leads to higher L2 proficiency, in interaction with age effects (R. M. DeKeyser, 2012). General biological mechanisms (e.g. gene shuffling) have been related to individual differences in language learning aptitude (Schumann, Crowell, Jones, Lee, & Schuchert, 2004). Individual differences in L2 learning success can also be related to motivational differences. For example, L2 learning success can depend on the strength of motivation to integrate in a host society or to fulfill pragmatic needs (Gardner & Lambert, 1972). Importantly, teaching and differences in the effectiveness of learning strategies (Cook, 2013) can also greatly affect learning success. In addition to individual factors, contextual factors affect L2 learnability across the L1s of individual leaners (Chiswick & Miller, 2005; Van der Slik, 2010; Van Tubergen & Kalmijn, 2005).

A three-way distinction often made in economic studies relates the factors involved in L2 learning to efficiency, exposure, and incentives (Beenstock, Chiswick, & Repetto, 2001; Chiswick & Miller, 2001). Efficiency is comprised of innate and learned cognitive capabilities that enable the learner to encode grammar, profit from language background, and allows for biological predispositions. A younger age and higher aptitude increase learning efficiency. Schooling and gender effects also affect learning efficiency although cultural differences play a role too (Chiswick & Miller, 1994, p. 19; Stevens, 1986; Van der Slik, 2010). An efficient language learner will hardly be successful in learning a target language without sufficient exposure to the target language, by means of either study hours or full immersion. Efficient learners who have a linguistically similar L1 and / or a higher education (e.g. writing skills) need fewer hours of study. Thirdly, exposure depends on the willingness and possibility to prioritize language learning. For example, parents may prefer to raise their children in the parents' mother tongue. Exposure costs time and money, and motivation is required to invest these resources.

Language background effects on L2 learning have been studied extensively in the context of cross-linguistic influence (Jarvis & Pavlenko, 2008; Kellerman & Sharwood Smith, 1986). For example, studies have investigated the problem that consonant clusters may impose on language learners. When adult learners are presented with an artificial language that contains unlikely consonant clusters with respect to their native language, the learners do not use information of unfamiliar consonant clusters (Boll-Avetisyan & Kager, 2013). In addition, a canonical example of language background effects in learning L2 phonology is the persistent problem that Japanese speakers have with the /l/-/r/ distinction in English, as this contrast does not exist in Japanese (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Flege, Takagi, & Mann, 1995; Sheldon & Strange, 1982).

How do learners deal with new L2 varieties? Possibly, L2 learners make use of the general learning mechanisms already available to them. The L2 learning task requires L2 learners to generalize from their L1 to a new L2 variety. Language background effects on L2 learnability provide an opportunity to study how learners deal with variability between languages, which is a salient property of the L2 learning task (e.g. bridging the gap). Providing further evidence of their linkage, the L1 and L2 learning processes themselves cannot be separated unambiguously as not only the L1 influences L2 processing, but also the L2 influences L1 processing. For example, the L2 has been shown to affect the expression of manner of motion in the L1 (A. Brown & Gullberg, 2008). Additionally, evidence from L3 learning shows the role of additional language background (the L2) besides the role of the L1 only (Amaro, Flynn, & Rothman, 2012; Bardel & Falk, 2007; Cenoz, Hufeisen, & Jessner, 2001; Hufeisen & Jessner, 2009).

Researchers have few opportunities to study variation in L2 learnability across multiple L1s. Psycholinguistic studies of cross-linguistic influence on adult L2 language acquisition typically employ an experimental group-design in which groups of learners have to acquire certain specific features (Ionin & Montrul, 2010; C. Lew-Williams & Fernald, 2010; Van Hell & Tanner, 2012). The inclusion of many learner groups in such designs is time-consuming. Without many groups, such designs do not allow insights into variation in L2 learnability across L1s.

Few studies compare L2 learnability across a multitude of groups. One study by the Foreign Services Institute of the US Department of State estimated the number of hours that English-speaking US citizens needed to learn the basics of a number of foreign languages. The researchers collected L2 proficiency scores (in the context of language classes) for many learners after 24 weeks of training. The aggregated levels of proficiency across the foreign language quantify L2 learning difficulty of a number of L2s for native speakers of English. The resulting L2 learning difficulty measures explained differences in US and Canada immigrant self-reported speaking proficiency score levels (Chiswick & Miller, 2005) and correlated strongly with linguistic distance measures (Cysouw, 2013).

Another source of speaking proficiency scores comes from language-testing data, which can potentially provide more sophisticated proficiency scores across more language backgrounds (Kim & Lee, 2010; Snow, 1998; Van der Slik, 2010). Potentially useful sources of language-testing data come from language-testing institutions such as the English language proficiency tests IELTS, CPE, and TOEFL. Participants in a language exam pass it after many weeks or sometimes years of studying. In contrast to self-reported measures of proficiency, language-testing data provide relatively objective measures of overall proficiency for learners with a wide range of different mother tongues. Language tests determine proficiency scores according to precisely established procedures and criteria. It is important that items are not biased for certain learner groups (Xi, 2010). Essentially, language-testing scores are correctness ratings made by second language teaching experts with a specialized training.

The state exam *Dutch as a Second Language* (STEX) is an example of such a useful source to study variation in adult L2 learnability. STEX consists of tests for speaking, writing, reading, and listening proficiency. To be successful, a candidate needs to pass all four parts of the exam. Compared to writing, reading and listening, speaking may be least affected by (advanced) educationally learned skills. However, language background effects also exist in writing, reading, and listening (Harding, 2012; Koda, 1989; Kubota, 1998; Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2009). The test outcome for the speaking part depends on approximately 60 content and form ratings of speech tasks to be performed by the candidates. The speaking exam does not test conversation as such (Levinson, 1979, 1980), but is limited to “one-directional” language. The speaking exam consists of about 30 tasks that vary in length and the number of speech tasks that the learner needs to accomplish. For the speaking proficiency exam, on which we focus in this thesis, grammar, vocabulary, and pronunciation need to conform to a specific standard. The ratings assess adequateness of content, wording, pronunciation, pace, vocabulary, register, coherence, and word order, amongst others. The passing criteria follow the criteria of the B2 level of the Common European Framework of Reference of Languages (Council of Europe, 2001). These criteria require learners to provide adequate reactions in a range of situations of daily life and business.

To summarize, L2 learning difficulty depends on a host of factors. Most likely, general learning mechanisms are involved in both L1 and L2 learning. Few studies have been able to generalize to a feature-general approach to study variation in adult learnability of additional language conditional on previously learned languages. Language-testing scores may provide a unique opportunity to tackle this problem. Language testing scores are available for many L1-L2 combinations. As these scores provide domain-general measures of proficiency, they may be useful to investigate what linguistic distance measures can explain differences in proficiency scores across L1s. We discuss various approaches to measure linguistic distance in the next section. These approaches include measures of distance in the lexical, morphological, and phonological domains.

Linguistic Distance

This section introduces what distance measures may explain variation in the learnability of an additional language. A consideration of many L1-L2 pairs may shed light on what dimensions or features of earlier acquired languages are important for language learning. We wanted to study variation in L2 learnability in terms of linguistic distance, as we think that distance is essential to general learning mechanisms involved in language learning and in learning tasks in general.

Languages vary widely in their lexical, morpho-syntactic and phonological make-up. Many comparative language and typological studies can be found in these three domains. Such studies try to define the differences and the similarities that can be used to define distances between languages, including distances between a wide variety of L1/L2 languages and Dutch as an additional language (L2/L3). Adult learners have to bridge gaps between the lexical, morphological, and phonological structure of their language background and the target language. The literature has different ideas of what linguistic gaps are and what linguistic distance measures reflect these gaps best. These ideas vary across the lexical, morphological, and phonological literature. The lexical literature discusses whether historical relationships and the degree of evolutionary change between languages model distance better than absolute measures of shared cognacy between the lexicons of a language pair. The morphological literature distinguishes between grammatical structures that are more morphologically distant from each other depending on their complexity. This way, a complex morphology is more distant to a simple morphology than vice versa, resulting in asymmetric distances. The phonological literature discusses whether differences between inventories of phoneme segments or inventories of distinctive features determine the differences between phonologies. In all, for the development of new distance-based models of L2 learnability, we start from the idea that a larger distance results in lower L2 learnability.

The three sections below introduce how lexical distance, morphological complexity, and phonological similarity may explain variation in L2 learnability.

Lexical Distance (Chapter 2)

Immigrant studies of L2 language proficiency have recently shown a renewed interest in explaining variation in proficiency using measures of linguistic distance as based on genetic relatedness (Isphording & Otten, 2011, 2013, 2014; Kim & Lee, 2010; Van der Slik, 2010). These studies show that linguistic distance as based on measures of genetic relatedness explain differences in L2 language proficiency. The effects are found among immigrants with various language backgrounds learning one target language, as well as in immigrants with mixed destinations and origins (Van Tubergen & Kalmijn, 2005). Most immigrant studies equate the language background of immigrants with the dominant language in the home country, although some researchers have also tried to disentangle country from language specific variation (Beenstock et al., 2001; Fearon, 2003). Studies of cross-linguistic influence and immigrant studies of target language proficiency may benefit from insights into the properties and principles underlying the linguistic distance effect. The most simple and very abstract measures of linguistic distance as based on genetic relatedness count the number of nodes that are shared between two languages (Adsera & Pytlikova, 2012; Desmet, Weber, & Ortuño-Ortín, 2009; Isphording & Otten, 2014; Van Tubergen & Kalmijn, 2005) in a language family classification (e.g. P. Lewis, Simons, & Fennig, 2013). Such measures assume that a closer genetic relatedness between two languages results in a closer linguistic distance. For example, German and English share the nodes Germanic and Indo-European while French and English only share Indo-European.

One potential way to improve such abstract measures is to use a measure of genetic relatedness that accounts for the degree of evolutionary change between two languages, or the amount of time since two languages branched off from each other (Bouckaert et al., 2012; Gray & Atkinson, 2003). Such estimates are available from phylogenetic language family trees (Forster & Renfrew, 2006). Currently, phylogenetic language family trees make use of sound recurrences in small lists of cognates (Swadesh, 1952); words with a shared common ancestor, to determine the most likely degree of evolutionary change between the branches of the tree. Family trees based on structural data are applicable as well (Dediu & Levinson, 2012; Dunn, Terrill, Reesink, Foley, & Levinson, 2005). Evolutionary constraints produce divergence in lexicons that allow reconstruction algorithms to figure out the genealogical structure in the family back to a certain point in time at which divergence estimates become too uncertain, called the family root. Currently, there may be around 7000 living languages that can be classified into 147 families (P. Lewis et al., 2013). Evolutionary constraints are drivers of such an immense diversity of language varieties (Gavin et al., 2013; Levinson & Gray, 2012). Besides phylogenetic trees as based on linguistic data, another approach compares the genetic make-up of the speakers themselves (Cavalli-Sforza, 1997). However, the speakers of a language and their genetic make-up may get blurred over time so they are not necessarily associated any longer with each other (O'Grady et al., 1989). As a result, a measure of linguistic distance based on genetic differences between speakers is inherently unreliable.

Cognacy measures are affected by comparable problems. Experts need a detailed account of the common recurrent sound correspondences between two languages to determine whether a translation pair shares a common ancestor or whether it is a chance resemblance or a result of borrowing and diffusion patterns (Thomason & Kaufman, 1988, p. 39). Some translation pairs have a similar form not because they retained the form over time, but because of borrowing and diffusion patterns. Diffusion of one word into an area where another language is spoken can happen when two languages get into contact (Dixon, 1997, p. 19). In addition, speakers can borrow words from a newly learned language into their native language. Diffusion and borrowing also contribute to linguistic similarities, alongside similarities due to genetic inheritance. For example, English shares more words with a similar form and meaning with French than with German or Dutch (Schepens, Dijkstra, & Grootjen, 2012), because of an extended period of language contact between French and English in the 11th century. In contrast to phylogenetic measures, which account for evolutionary change, form similarity measures measure the degree of disparity between languages.

It is currently unclear whether a measure that is based on genetic relatedness as well as on borrowed words captures distance effects on L2 learnability better than measures that are based on genetic relatedness or patterns of borrowing and transfer alone. A variety of linguistic distance measures exist that express the form similarity between words from two different lexicons in a numerical way (Heeringa, 2004; Kondrak, 2000, 2001; Kondrak & Sherif, 2006; McMahon & McMahon, 2005). An important aspect of learning an L2, besides grammar and pronunciation, is the expression of meaning with words in an L2. We use the term lexical distance when a linguistic distance measure is based on word form comparisons or recurrent sound correspondences in words. Chapter 2 compares the explanatory value of two lexical distance measures with each other. The first measure is based on language diversity as measured by the differences in the degree of evolutionary change between languages (Gray & Atkinson, 2003). The second measure is based on language disparity as measured by the proportions of words with similar form between languages (C. H. Brown, Holman, Wichmann, & Velupillai, 2008; Holman et al., 2008).

Chapter 2 introduces an approach for comparing different linguistic distance measures with each other. The approach uses language testing speaking proficiency scores to compare their explanatory value. We use multilevel models to decompose variance in speaking proficiency scores into an individual learner level variance component, a contextual language level component, and a contextual country level component. Contextual components are only observable after aggregating over individual learners. Multilevel models enable the estimation of these variance components while simultaneously regressing proficiency on individual and contextual level factors. The linguistic distance between the mother tongue and the second language (L2) of a learner is a contextual effect that varies according to the degree of difference between a learner's mother tongue and the L2. We study interactions between linguistic distance and the quality of the educational system, years of full-time education, gender, length of residence, age of arrival in the Netherlands, and proficiency in an additional language. Factors besides linguistic distance can potentially blur the effect of linguistic distance, which makes it important to study whether effects of linguistic distance are robust against effects of third factors.

Morphological Complexity (Chapter 3)

Adult language learners seem to experience great difficulty in learning the derivational and inflectional morphology of an additional language. Is learning a morphological complex language generally difficult or do more problems occur when an L2 is morphologically more complex than the L1 of a learner? Typologists (Dahl, 2004) and sociolinguists (Trudgill, 2011), making use of historical evidence, define morphological complexity relative to or in relation to L2 learning difficulty. Features of language that are morphologically complex can be identified by comparing language of L1 speakers with interlanguage of language learners (Selinker, 1972). Interlanguage is more likely to include features that are easy to learn and exclude features that are difficult to learn (Kortmann & Szmrecsanyi, 2012; Szmrecsanyi & Kortmann, 2009). For example, when beginning Italian learners of German want to focus on a subject, they prefer to employ word order instead of inflectional morphology, e.g. "mädchen nehme brot" (W. Klein & Perdue, 1997).

It has recently become possible to make typological comparisons across many languages simultaneously to evaluate differences in morphological complexity (Bentz, Verkerk, Kiela, Hill, & Buttery, Submitted; Bentz & Winter, 2013; Dale & Lupyan, 2012; Lupyan & Dale, 2010). Large-scale typological databases such as WALS enable the development of linguistic distance measures based on differences in morphological complexity between the L1 and the L2 of a learner. An essential virtue of the use of typological data is that it overcomes the problem typically related to lexical comparisons, namely that lexical comparisons are limited to language combinations from one particular language family. We relate measures of morphological complexity to empirical measures of learning difficulty in Chapter 3. Chapter 3 investigates to what extent differences in morphological complexity have an impact on proficiency levels attained in L2 Dutch. We correlate typologically defined morphological distances between 49 L1s and L2 Dutch with variation in L2 learnability. We investigate a previously analyzed set of 28 morphological features (Lupyan & Dale, 2010) to study both correlations for individual features differences in complexity and correlations with an overall measure of feature complexity differences. We also study whether a decrease in morphological complexity correlates with variation in L2 learnability in order to find out whether linguistic differences that involve a step up in complexity are more difficult to learn than linguistic differences that involve a step down in complexity.

One of the currently outstanding issues is the importance of typological compatibility between languages for the facilitation of structural borrowing and transfer (Aikhenvald & Dixon, 2006). It is not clear what the consequences of two incompatible (linguistically distant) language structures are for patterns of borrowing and transfer in a contact situation. A popular explanation of language simplification (e.g. Trudgill, 2001) is that a merger of language communities in terms of a high incidence of L2 speaking in particular decreases the morphological complexity of a language (adaptation). Language adaptation is a process of linguistic change directed by social changes or innovation in the speaking population of a language (Levinson & Gray, 2012). For example, new words will be introduced with the invention of new technologies such as dye and paint and more complex sentences can develop with development of more advanced written language (Karlsson, 2007; Levinson, 2000).

Chapter 3 speaks to this issue by providing empirical data to evaluate the effects of morphological complexity on L2 learnability, with the aim of providing new insights into the underlying reasons why L2 speaking can decrease the morphological complexity of a language. As for now, some empirical studies exist that provide correlational evidence for the link between adaptation and L2 learning biases (Bentz et al., Submitted; Bentz & Winter, 2013; Dale & Lupyan, 2012; Lupyan & Dale, 2010). However, it is still unclear how the link depends on L1-L2 differences. Chapter 3 shows the role of L1-L2 complexity differences in L2 learnability across a large range of L1s with varying degrees of morphological complexity.

Phonological Distance (Chapter 4)

The new sounds in the phoneme inventory of an L2 lead to pronunciation problems that even persist throughout life for most learners of an L2 (Flege, Munro, & MacKay, 1995; Munro, 2008; Piske et al., 2001). Only in exceptional cases do adult learners attain a native-like level as an adult learner (Bongaerts et al., 1997). The articulatory movements necessary for the production of L1 sounds seem to occur without much conscious effort. Does the structure of the L1 phoneme inventory influence the learnability of an L2 phoneme inventory? In some sense, an L2 learner needs to rewire the movements of the articulators to accommodate the movements necessary to produce new sounds and new combinations of sounds. To determine the effect of phonological similarity on L2 learnability, it may thus be necessary to compare the sound inventories in relation to the way they are articulated (C. Brown, 1998) and not at the level of the sounds themselves. It is unclear however, what distinctive features between phoneme inventories at the level of the articulators contribute to L2 learnability.

The learner of an L2 phoneme inventory needs to learn the speech sounds of a target L2 that are not already part of the L1 sound inventory. Three categories of problems in learning new sounds can arise. First, a learner can accidentally substitute a new sound with a similar L1 sound. Second, a learner can fail to perceive all the phonetic detail of a new sound. Third, a learner can fail to learn L2 phonotactic constraints (Major, 2008). New sounds can range from almost similar to completely different compared to the articulatory structure of L1 sounds. Analogous to the idea that a higher distance results in lower L2 learnability, we hypothesize that learning the complementary features of new sounds results in lower L2 learnability. Moving beyond the level of the number of new sounds in phoneme inventories to the level of distance to new sounds in terms of distinctive features may provide the necessary units of comparison to test whether phonological distance successfully explains variation in L2 proficiency scores (C. Brown, 1998; McAllister, Flege, & Piske, 2002; Michaels, 1974; Ritchie, 1968).

The core aim of Chapter 4 is to test if the difficulty of learning the phonology of an L2 depends on the number of new sounds and / or on feature-based similarities of the new sounds compared to the sounds of the L1. We test the relationship between L2 learnability and sound inventory level measures as well as feature level measures of similarity between sounds. The feature level measures distinguish between symmetric and asymmetric overlap of the features of sounds, difference and similarity, and presence and absence of features.

Variation in Adult L3 Learnability across L1s and L2s

In the previous three sections on linguistic distances, we introduced three studies that each question what specific instantiations of L1-L2 linguistic distance measures may explain variation in L2 learnability in the lexical, morphological, and phonological domains. In addition to evaluating these measures on variation in *L2* learnability, they may also explain variation as L2-L3 distance measures in an *L3* learnability approach. As L2 learnability depends on the L1, L3 learnability depends on both the L1 and the L2.

L3 learning is a widespread phenomenon. Only a small group of learners of Dutch (less than 20%) does not speak an additional language besides their L1 when they take the L2 Dutch state exam (based on STEX data, see Chapter 5 and 6). It is theoretically important from a societal as well as a linguistic point of view to understand how effects of an additional language background compare to effects of the first language. The two sections below introduce two studies that analyze variation in L3 learnability depending on the L1 and L2. The first study investigates how to decompose variation in L3 Dutch speaking proficiency scores into by-L1 and by-L2 variation. The second study investigates whether both L1 and L2 lexical and morphological distance measures are successful models in predicting the problems that adult multilingual learners encounter when learning Dutch as an additional language.

L3 Learning (Chapter 5)

Variance in language testing scores poses statistical challenges for the investigation of its various underlying sources of variance. We want to compare the relative contribution of speaking language x as an

L1, and speaking language y as an L2, controlling for other languages, other countries, and third factors such as age and education. A specific statistical toolbox called multilevel regression that allows for such inferences has been developed over the past decades (Goldstein & McDonald, 1988; Raudenbush, 1993). Multilevel analysis is currently finding its way to language research (Baayen, 2008), although its implementation is still being actively optimized and improved, especially for large-scale data (such as STEX) with partially crossed random effects (D. Bates, Mächler, Bolker, & Walker, 2014). Multilevel models decompose variance in a dependent variable into separate variance components by means of computationally heavy integration over the variables in the model. The resulting estimates of variance components are probability distributions over the levels of a random effect (subjects, items, languages, countries, etc.) that allow for the inference of most optimal adjustments for each level of a random effect, which are called best linear unbiased predictors (BLUPs) (G. Robinson, 1991). BLUPs are useful for comparing the relative contribution of different levels of random effects, i.e. individual languages. What are the assumptions of the BLUPS and how do modeling decisions affect the estimation of BLUPs?

Chapter 5 investigates the effects of different random effect structures in multilevel analysis on the effects of L2 distance on L3 learnability. For example, variation in L2 speaking proficiency results from by-learner, by-teacher, and by-school variance, which are three hierarchically organized random effects. A subclass of multilevel models deals with cross-classified random effects. This class is often called mixed effects models. For example, the mixed effects analysis of many linguistic experiments treats by-subject and by-item variation as two crossed random effects. Mixed effect models, or more specifically cross-classified random effect models (CCREMs), always assume that multiple random effects are independent. However, crossed random effects may be interdependent to some extent. The consequences of such an assumption are currently ill understood. We want to investigate the consequences of this assumption by comparing L1 and L2 influences on proficiency in Dutch as an L3. Using STEX, we aim to assess the mutual dependency between the crossed random effects of the L1 and L2 on L3 learnability. In particular, Chapter 5 investigates whether the variation across L1s and L2s is comparable and whether the estimation of this variation depends on assuming that the L1 and L2 effects are independent.

The L1 and L2 Distance Effects (Chapter 6)

The L3 literature has proposed at least three explanations of how the L1 and L2 interact when learning an L3. First, the background language with the lowest distance will have most influence (Rothman, 2011). This means that learners will transfer from the language they believe is most typologically similar to the target language. Second, the more recently the language is learned (which is the L2), the more it will block previously acquired languages (Bardel & Falk, 2007; Bohnacker, 2006). Third, the L2 plays a role that is either neutral or positive, and it is more beneficial for learning an L3 than having no L2 at all (Flynn, Foley, & Vinnitskaya, 2004). In order to evaluate how L2 distance interacts with L1 distance, we want to evaluate these propositions on the STEX data. In addition, we investigate the relative importance of L1 and L2 lexical and morphological distances, and how learners combine L1 and L2 distances.

Besides the multiple linguistic sources available to the learner of a third language, learners may develop an abstract multilingual awareness and a set of skills that allows them to make faster and more accurate inferences (Jessner, 2014). If multilingual learners of Dutch reach higher average performance than monolingual learners of Dutch, do the specific linguistic distances involved explain the benefit of the multilingual condition, or is there still a benefit after distance has been accounted for?

As we described above, Chapters 2 and 3 test whether lexical and morphological distance successfully explain variation in L2 learnability. Having decomposed variation into L1 and L2 variance components in Chapter 5, Chapter 6 tests whether lexical and morphological distances from Chapters 2 and 3 successfully explain variation in L3 learnability across L2 learners as well. A discussion of future work on L2 phonological distance effects is also included.

Summary of Objectives

In all, this thesis aims to show effects of linguistic distances on adult learnability of Dutch as an additional language in the lexical, morphological, and phonological domains. Chapter 2 tests lexical distance effects on adult L2 learnability across Indo-European L1s and its robustness against interactions with third factors such as age and exposure. Chapter 3 investigates the effects of morphological complexity on adult L2 learnability across both Indo-European L1s and non-Indo-European L1s. In particular, chapter 3 aims to show that learning additional morphological complex features is relatively hard and why such insights from L2 learnability are important for the study of cultural-evolutionary mechanisms in language divergence. Chapter 4 studies phonological similarity effects on adult L2 learnability across both Indo-European and non-Indo-European L1s. In particular, chapter 4 aims to show that the distinctive features of new L2 sounds influence L2 learnability. Chapter 5 and 6 focus on Dutch as a third language. Not only L1 distance effects are included as predictors, but the strongest earlier acquired L2 (if present) is included as well. Chapter 5 tests whether the relative ordering of by-L2 adjustments is comparable to the relative ordering of by-L1 in predicting proficiency in Dutch as an additional language (L3) and whether this ordering is robust against L1-L2 interactions in the large set of L1 and L2 combinations. The objective of Chapter 6 is to test if lexical and morphological distance affects adult L3 learnability across both Indo-European and non-Indo-European L2s. In addition, it investigates whether L2 distance effects are additive to L1 distance effects in predicting L3 Dutch proficiency. In all, Chapters 2, 3, 4 test whether typological distance effects provide accurate models of the difficulties that adult learners have when acquiring a second language and Chapters 5 and 6 study whether these distance effects also pertain to other previously acquired languages besides the first language, in particular the strongest L2, if present. All five chapters make use of the STEX data that we describe next. These chapters have been designed as independent publications. The beginning paragraphs of the methods sections of each of these chapters offer specifics about STEX. These beginning paragraphs are not required for readers who are reading this dissertation from cover to cover.

A Description of STEX

We used a unique database of state exam L2 proficiency testing scores (STEX) for our studies of linguistic distance and adult L2 acquisition. The secretary of the board of Dutch state exams made the (anonymized) data of the period 1995-2010 available, amounting to proficiency scores for more than 50,000 learners. The board of state exams (currently called College voor Toetsen en Examens) is the liable owner of the data. The Dutch government installed an advisory committee to develop STEX in 1991 and implemented STEX in the following year. Although formally not related to each other, STEX superseded and replaced a test that Dutch universities had been using. Successfully passing STEX provides candidates with a stepping-stone to access the education and labor market, according to the official committee that developed the state exam. Meeting integration requirements is not part of the purpose of STEX, as language assessment became part of the integration requirements for immigrants after STEX had been developed. However, immigrants can decide to do the STEX exam, which offers them a higher proficiency level than strictly necessary to pass integration requirements.

This section discusses the test design, the independent variables, and selection steps, in order to give a general overview of STEX and to present details that are not included in the following chapters. For further descriptions of the languages used in our studies, we include aggregated speaking proficiency scores summarized by L1 and L2 in Appendix A.

Test Design

STEX tests the overall level of speaking, writing, reading, and listening proficiency. The Centraal Instituut Toetsontwikkeling (CITO) and the Bureau Interculturele Evaluatie (ICE), two large test battery constructors in The Netherlands, jointly construct the exam and the exam questions. Over the 15 years of testing, the testing design has been constant, meaning that the structure of the speaking exam from 2010 closely resembles the structure of the speaking exam from 1995. The speaking exams consisted of about 14 tasks that vary in length. In longer tasks, candidates needed to give a detailed opinion, argumentation, or description. For example, candidates had to respond to the question “In Dutch television a lot of ads are made for all kinds of products, even in the middle of a program. What is your opinion about ads on TV?” The speaking exams took about 30 minutes. Candidates received detailed instructions through headphones. Subsequently, candidates had to give oral responses to the tasks, which were recorded on tape. The tasks required the candidate to produce different speech acts. For example, provide information, give instructions, congratulate, refuse, complain, apologize, state an opinion, tell a story, and so on. The use of dictionaries was prohibited. Two experienced examiners evaluated each recorded response independently on the basis of a list of detailed and specific criteria. In case of disagreement, a third examiner was called in.

Language production was assessed with respect to the functional, communicative language proficiency of the candidates. This implies that intelligibility is more important than e.g. a foreign accent. The actual content itself did not have to be correct (e.g. the names of Dutch TV channels), although the utterances had to be comprehensible. The candidate’s score was the average of the ratings assigned by the two examiners (NT2 State Examination, 2008).

The difficulty of the examinations was kept constant over time by applying a specific item response theory model, i.e. the one-parameter logistic model, which is an advanced Rasch model (Verhelst & Glas, 1995). Item response theory is generally used in large-scale assessment programs (such as those administered by the Educational Testing Service) to estimate candidate performance on the same scale while controlling for differences across items. Rasch models solve this problem by requiring that comparison between candidate ability is independent from item difficulty and vice versa. A decisive advantage of item response theory, as compared to models based on classical test theory, is that the test scores of candidates are allocated to the same ability distribution, even when they took different versions of the exam; hence, their test results can be analyzed simultaneously. The scores on the exam were standardized: 500 points or more implied that the candidate had passed the exam. The assessment criteria are comparable to the criteria of the B2 level (i.e., upper-intermediate level) as defined in the Common European Framework, which is comparable to a band score of 5.5 in the International English Testing System (IELTS).

Variables

The STEX data contains variables obtained through the candidate administration procedure, responses from questionnaires completed on a voluntarily basis, and the test results themselves. The administrative variables include gender, date of birth, and date of the exam. The questionnaire contained the following questions, translated from the Dutch originals:

- Since when have you lived in the Netherlands?
- What is your country of birth?
- What is your mother tongue?
- Do you speak an additional language besides Dutch and your mother tongue?
- If yes, which additional language? If you speak more additional languages, name the language that you know best.
- How many years of full-time education did you have?

The questionnaires also included questions that are not relevant to discuss here as none of the following chapter makes use of them. These other questions concerned the hours of lessons in Dutch as an L2 (Van der Slik, 2010), whether the candidate's motivation was to fulfill educational requirements or integration requirements, and whether a candidate works or studies as a full-time occupation. Over the period of 1995-2010, some of the questions changed in wording. For example, before 2005, participants in the exam could answer the question about full-day education by indicating a specific number of years, while after 2005 the answers were limited to 0-5, 6-10, 11-15, 16 or more years.

Selection Steps

Four tests are required to pass the exam in Dutch as an L2. We chose to focus on the speaking proficiency test only. Before 2005, candidates could take the speaking proficiency test only twice a year. After 2005, the number of exam moments was increased to 30 per year. The number of candidates in a specific year also varies due to the variation in the rate of immigration. Some candidates took more than one speaking exam, in case they did not pass their first exam. We chose to analyze only the first exam scores of the candidates. As noted above, the difficulty of the exam was kept constant over time, which ensures comparability of speaking proficiency scores from exams taken at different moments.

The exam requires a significant investment of time and money by the candidate. However, this does not mean that every registration leads to an examination. One of the reasons for missing data in STEX is that candidates do not show up. Another reason for missing data in STEX is a missing questionnaire or missing answers to the questionnaire. We removed all candidates with missing answers on age of arrival, the country of birth, mother tongue, and additional language background questions. The language coding used in STEX is STEX-specific and Dutch. We translated the coding into English and added ISO codes using both Ethnologue (P. Lewis et al., 2013) and WALS (Dryer & Haspelmath, 2011). We interpolated missing values for the full-day education question using the average values for all candidates from the same country of birth (889 cases). For all chapters, we removed all countries of birth, mother tongues, and additional languages with less than 15 candidates. The resulting data consists of speaking proficiency scores of 50,236 learners of L2 Dutch.

Chapter 2

The Effect of Linguistic Distance across Indo-European Mother Tongues on Learning Dutch as a Second Language

Acknowledgements
We would like to thank the secretary of the State Examination of Dutch as a Second Language for providing the learner data used in this chapter. We also thank our anonymous reviewers, Søren Wichmann and Quentin Atkinson for providing linguistic data, and Theo Bongaerts, Martin Becker, Malte Viebahn, and Thordis Neger for comments on an earlier version of this manuscript.

This chapter was published as:
Schepens, J.J., Van der Slik, F.W.P. & Van Hout, R.W.N.M. (2013). The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In L. Borin & A. Saxena (Eds.), *Comparing Approaches to Measuring Linguistic Differences (Trends in Linguistics. Studies and Monographs (TiLSM0, 265)*, pp. 199-230. Berlin: De Gruyter Mouton.
http://www.degruyter.com/viewbooktoc/product/203901

Abstract

Using multilevel models, we decomposed variance in proficiency scores across adult learners of Dutch as a second language into individual (learner characteristics) and contextual (group characteristics) components. The linguistic distance between the mother tongue and the second language (L2) of a learner is a contextual effect that varies according to the degree of difference between a learner's mother tongue and the L2. We have analysed L2 learners' state exams for Dutch speaking proficiency to explain variance in L2 proficiency scores on the basis of two different linguistic distance techniques: one that uses the traditional expert-based, historical-comparative method as input for its Bayesian phylogenetic inferences as used in Gray and Atkinson (2003), and one that uses automatic distance based method applied in the ASJP project as input for a neighbour joining algorithm (C. H. Brown et al., 2008).

We used data from more than 33,000 examinees, speaking 35 different Indo-European languages, originating from 89 different countries. Our main aim was to partial out the impact of linguistic distance on proficiency in speaking Dutch as an L2. The multilevel models that we used incorporated one confounding variable on the contextual level: the quality of the educational system, and five confounding variables on the individual level: gender, educational level, length of residence, age of arrival in the Netherlands, and proficiency in an additional language. We were able to identify robust L1 distance effects, for both of the distance measures we used, and we compared them to the predicted scores obtained in our multilevel analysis. Our conclusion is that differences in second language learning proficiency offer an excellent testing ground not only for validating the concept of linguistic distance itself, but also for comparing the performances of different types of linguistic distance.

The Effect of Linguistic Distance across Indo-European Mother Tongues on Learning Dutch as a Second Language

Introduction

It is a commonplace to state that learning a mother tongue (L1) is successful in most circumstances, but that learning a second language (L2) returns a less evident result. L2 learners diverge widely in their degree of success in acquiring a new language. The central question here is whether linguistic distance measures between the L1 and L2 are suitable instruments to predict the degree of success in learning an L2. The assumption is that the larger the distance the harder it is to learn another language. Establishing a clear relationship between L2 learning and linguistic distance gives strong support to external validity of the concept of linguistic distance.

Where do language similarities and dissimilarities come from? Looking back in history, one can see how languages diverge and converge. The Austronesian expansion of settlers to unexplored Polynesian islands established divergence step by step, causing new innovations to appear in a clear tree-like fashion (Gray & Jordan, 2000). In contrast, in the Russian Empire, language convergence by standardization was a crucial tool for excluding other languages and language variation (Ostler, 2005).

Processes of divergence and convergence have led to a complex distribution of many languages over many countries in the world. However, many countries explicitly opt for one single standard language in their language policy. As a consequence of massive migration waves, large groups of adults need to learn the (standard) language of the country of immigration. Tests and exams have been developed to test their L2 proficiency levels. In the Netherlands, for example, most immigrants have to pass the official state exam called “Dutch as a second language”.

In a previous study, a substantial amount of between mother tongue variance was explained with a measure of linguistic distance between the L1 and L2 on the basis of 11 West-European languages (Van der Slik, 2010). In the present study, we want to deepen our understanding of how barriers in learning an L2 are related to linguistic distances. We do so by expanding the set of L1 languages to all Indo-European languages (35 in our database), spoken in different countries (89 in our database), and by testing two different linguistic distance measures.

The remainder of this introductory section contains a discussion of current approaches in measuring linguistic distances, the effects of linguistic distance on L2 learning, and the approach taken in the present chapter.

Background

Approaches to Measuring Linguistic Distance

Recent discoveries in the dynamics of linguistic change disclose lineage dependent structural relationships in the evolution of word order in three large language families (Dunn, Greenhill, Levinson, & Gray, 2011). By reconstructing language family trees, it was shown that certain states of development are more likely given a previous state of development in a particular language family. The study of Dunn et al. is a recent example of a quantitative diachronic approach in which tree-like phylogenetic models are applied to language variation and change. Phylogenetic analysis uses the finding that linguistic data contain deep historic signals that can be used to date language branching (Crystal, 1987).

The treelike model of language evolution can be inferred and reconstructed from lexical (Gray & Atkinson, 2003), (morpho-) syntactic/structural (Dunn et al., 2005), or phonological data (Atkinson, 2011). Each of these three data types has its own limitations. The lexicostatistical approach, based on the comparative method to estimate cognacy, is an early method for inferring language relatedness; the structural and phonological approaches are fairly recent. Linguistic comparison on the basis of each of these different data types may produce different linguistic distance measures. In this paper we will apply lexical distance measures, although in the near future we intend to expand our research to the phonological and (morpho-) syntactical domains of linguistic distance.

The dominant lexical distance measures are based on the percentage of shared cognates between languages. Cognates are words that historically relate to the same word in a common ancestor language. Cognates can share form and meaning, just like borrowings and accidental form resemblances, which do not have a shared origin. Cognacy can be qualitatively coded, as in the comparative method (Dyen, Kruskal, & Black, 1992; Swadesh, 1952), or as a quantified degree of distance from one form to another (Heeringa, Kleiweg, Gooskens, & Nerbonne, 2006; Kessler, 2005; McMahon & McMahon, 2005). The distance-based method is based on the observation that cognates tend to share their form across languages, although not always in identical form. In the distance based method, string distances between two word forms can be automatically simulated. To exclude borrowing effects on measuring distance as much as possible, both the comparative method and the distance-based method are usually applied to (subsets of) Swadesh lists (C. H. Brown et al., 2008; Holman et al., 2008), which should sample from basic vocabulary. The percentage of shared cognates, or the average distance between words on the list, generalizes to a measure of linguistic distance between languages.

We used the linguistic distances found in two lexicostatistical studies. The first study (Gray & Atkinson, 2003) determined shared cognates on historical-comparative grounds in a binary way, the second study (Holman et al, 2008) determined the degree of cognacy of word pairs by computing string distances. Both studies carried out a phylogenetic analysis in order to retrieve the optimal tree-like structure from the distances obtained. In this chapter, we refer to the measurements of Gray & Atkinson (2003) as G&A, and to measurements described in Brown et al. (2008) and Holman et al.

(2008) as ASJP (the name of the project: the Automated Similarity Judgment Program).

The historical-comparative method entails a judgment process carried out by experts who are able to identify how sounds are preserved or have changed over time. Gray & Atkinson (2003) used expert cognacy judgments from Dyen, Kruskal, and Black (1992) and applied a phylogenetic analysis while imposing certain time-constraints on the tree-like structure. They retrieved the historical signals proportional to evolutionary change, including dates of linguistic innovations.

The similarity measures from the ASJP were computed automatically using a distance based method (Brown et al., 2008). For the distances used in this chapter, we used ASJP database version 13 (Wichmann, Müller, et al., 2010) and software from Holman (Holman, 2010, 2011) which computes normalized Levenshtein distance measures. Wichmann, Holman, et al. (2010) evaluated the normalization of Levenshtein distances by word length and average chance similarity.

ASJP-based linguistic distances can either be extracted as the average normalized string edit distance between the Swadesh lists of two languages, or as branch lengths from the resulting phylogenetic tree as computed using a neighbour joining algorithm (the correlation between the two distances is .986). The normalized string edit distance is the Levenshtein distance measure normalized by dividing it by its theoretical maximum (length of longest word). In ASJP, it is additionally corrected for chance similarity by dividing it by the average distance of words not referring to the same concept in that language pair. The measure was developed to be able to distinguish between related and unrelated language pairs.

As the ASJP automates the expert-based comparative method and G&A does not, we refer to the ASJP method as automated, and we refer to G&A as expert-based. However, there are other differences between the two methods as well. For example, ASJP categorically reduces sound inventories to a subset of possible sounds. The method and results section describe the differences between the automatic method and the expert-based method after applying them to the Indo-European languages from our dataset.

Second Language Learning Effects of Linguistic Distances

The best known predictor for transfer in second language acquisition is the degree of congruence between the source language (L1) and the recipient language (L2) (Jarvis & Pavlenko, 2008, p. 176; Kellerman, 1979). This constraint has been labelled “language distance”, “typological proximity”, “psychotypology” (perceived proximity), or “cross-linguistic similarity”. The effect of the mother tongue on second language learning was amply discussed within Contrastive Analysis (Lado, 1957; Odlin, 1989; Weinreich, 1963), but this method was not developed to determine or calculate linguistic distances.

The empirically based model proposed by Chiswick and Miller (2005, 2007) poses that language proficiency scores result from the interaction between incentives (motivation, money, labour), exposure (time, lessons), and capacity (education, talent, language background). Recent immigrant studies (Chiswick & Miller, 2005; Van Tubergen & Kalmijn, 2009) have found support for this model.

An important part of the effect of language background in the learner is determined by the effect of linguistic distance from one’s mother tongue to a destination language (Espenshade & Fu, 1997). Recently, the L2 effect was modelled with multilevel models by incorporating linguistic distance from learners’ mother tongues to Dutch on a contextual level (Van der Slik, 2010), using linguistic distance measures from McMahon & McMahon (2005). The effect of linguistic distance on second language proficiency of immigrants has been incorporated in only a few other studies, although mostly in a reverse way. In such a reverse approach, immigrant proficiency scores are explained by incorporating measures based on the ease or difficulty American emigrants experience in learning a specific language (Chiswick & Miller, 2005, 2007; Van Tubergen & Kalmijn, 2009). Such empirically determined differences in second language learning were also used to infer which typological features may be involved in second language leaning (Cysouw, 2013). This approach, in which the difficulty of learning a foreign language is accounted for, is problematic for various reasons. Most importantly, motivation among emigrants is expected to differ for different languages. A measure of linguistic distance from one's mother tongue to Dutch does not suffer from these impairments.

Recent studies also relate immigrants' proficiency scores to a quantified measure of linguistic distance (Isphording & Otten, 2011, 2013, 2014) by assuming that linguistic distance varies across migrants coming to Germany and the US. In addition, effect of linguistic distance may need to be explained across mother tongues. Neglecting this hierarchical structure may lead to an underestimation of standard errors and hence to a potential unjustified rejection of null hypotheses (Snijders & Bosker, 2011).

Most sociological and economic studies of language proficiency measure proficiency using self-report. However, this is not a valid way of measuring language proficiency as speakers tend to overestimate or underestimate their proficiency (Charette & Meng, 1994; Finnie & Meng, 2005). Immigrants may evaluate their skills relative to those of other immigrants rather than native level proficiency. Formal assessment by language tests overcomes these shortcomings of self-reports.

In our model, we incorporated quantified linguistic distance measures (G&A, ASJP) to explain the variance in scores on the state exam "Dutch as a Second Language". These measures may explain part of the variation in individual proficiency levels, together with other predictors. Overlap between linguistic and empirical measures may show why a high level of proficiency in Dutch is more easily attainable for some learners than for others on the basis of linguistic differences. In this way, linguistic distance could turn out to be an important but underspecified contextual factor in understanding learning differences in second language acquisition.

Empirical measures of linguistic distance need to take into account other contextual and individual differences that may affect performance on tests of L2 proficiency. This implies that a distinction has to be made between contextual effects such as linguistic distance and quality of schooling in the country of origin on the one hand, and individual effects such as length of residence on the other. Given the many multilingual countries in the world, identifying the effect of linguistic distance implies the necessity of separating on the contextual level the effect of the L1-L2 distance from country effects. For example, the country's estimated schooling quality for immigrants speaking Kurdish as their mother tongue can be the schooling quality of Turkey, Iraq, or of a number of other countries. Beenstock et al. (2001) also made a distinction between languages and country of origin, as they also tested linguistic distance by separating it from national characteristics.

Present Study

In a previous study on the Dutch state exam results, Van der Slik (2010) traced back the overall variation in oral and written proficiency in Dutch to a cognate (McMahon & McMahon, 2005) and a genetic measure (Cavalli-Sforza, Menozzi, & Piazza, 1994) to establish linguistic distances from eleven Western European languages to Dutch. The genetic linguistic distance, based on genetic differences between populations, explained less variance in language proficiency as compared to the cognate measure of linguistic distance.

In the present study we aim to extend the previous study in several ways. First, we apply two distance measures, one based on the expert-based traditional historical-comparative method (G&A) and one based on a gradual, automated measure (ASJP). These two methods allowed us to expand the number of L1 languages from 11 to 35 Indo-European languages in our analyses. We took a larger list of mother tongues to show that a linguistic distance based model is generally applicable to explain second language proficiency in speakers with different mother tongues.

The fact that the Indo-European language family is well-studied and that linguistic distance measures are relatively readily available for this language family is a persuasive argument to include all IE languages present in our dataset. This selection resulted in 35 different languages with speakers from 89 different countries.

We analysed test scores of more than 33,000 learners that took part in the Dutch language exam and we used exam scores from 15 years of immigrant history (1995 – 2010).

At the individual level of the learner, we used the model of language proficiency used by Chiswick and Miller to distinguish between indicators of capacity (measured by gender, years of full-time education), exposure (measured by age of arrival, length of residence), and incentives. Unfortunately, we had no measures on incentives at our disposal. Effects of capacity, exposure, and incentives may differ across learners according to other individual and contextual characteristics. For example, it is has been argued that less memory capacity is available for language learning at a higher age (Birdsong, 2014; Ullman, 2005). Therefore, learning a more distant target language might be more problematic for older learners, as more cognitive capacity is required than for learning a more similar language. The full disentanglement of such interaction effects is not the focus of this study, but we do hypothesize a pervasive presence of linguistic distance effects in different processes involved in language learning.

Language and country characteristics refer to distinct but related constructs at different contextual levels. Linguistic distance is part of the construct language characteristics. Given that languages are different, a quantified distance measure might explain effects of linguistic differences. Country of origin characteristics may include educational quality amongst others. Given that countries have organized their educational systems in quite different ways, we expect effects of quality of education on second language learning as well.

Educational quality is part of the construct country characteristics. We will focus on oral proficiency as the dependent variable.

Methods

We analysed test scores of Dutch language proficiency from the State Examination Board of Dutch as a Second Language (NT2), which is based on the Common European Framework of Reference for Languages. The NT2 exam scores were kept comparable over a time period of 16 years (1995 – 2010), with different spacing and structuring of tests each year, using an item response theory model. A proficiency level of 500 or more in all four different proficiency components (reading, writing, listening, speaking) determined exam success. 77.7% of all examinees in our dataset passed the exam at their first attempt. Participants were given the opportunity to register for as many exams as needed to pass all four components, but we took only test scores of first attempts. Participants were given the choice of taking exams specially tailored towards higher education (called STEX II; required for admittance to a Dutch university) or of taking exams for vocational training (called STEX I). Only scores on the STEX II exam were used in this study.

Sample

We selected all Indo-European languages with more than 30 speakers in our database in order to have a sufficient number to include context characteristics. The number of languages was 35, with 945 speakers per language on average (SD=1260). The selected languages were spoken in 89 different countries (at least 20 speakers per language per country, Mean=376, SD= 735). Combining languages and countries resulted in 118 groups, see the Appendix. The sample included test scores of 33,066 learners with an Indo-European mother tongue over a time frame of 16 years (1995 to 2010). 73% of the participants were women. We only included participants who answered a question on years of full-time education in the questionnaire that was given to participants, prior to the start of the exam.

Dependent Variable

Examinees had to perform different speaking tasks for 30 minutes. Performance was judged according to a formal judgment model on content, correctness, wording, pronunciation, pace, vocabulary, register, coherence, and word order, amongst others. Both the test and a formal judging scheme were jointly developed by CITO (central institute for test development) and CvE (board of exams). Both are Dutch institutions that develop and maintain large test batteries. A more detailed discussion of the language test can be found in Van der Slik (2010).

The exams took place at specific exam dates. Until 2005, there were four exam sessions a year, while from 2005 onwards there were 30 sessions per year. To pass the full STEX II exam, participants had to complete tests of listening, reading, speaking, and writing. Participants could choose to do different exams at different exam dates; therefore measurement points are generally not comparable across individuals. Furthermore, some individuals only participated in one, two, or three out of the four exams. Generally, doing the exams required a considerable amount of effort from the learner, both in training, as well as in arranging the different sessions for the four tests.

Contextual Characteristics

We defined the contextual level not only by language but also by country of origin in order to capture the intertwining of language and country characteristics and their cross-classifications.

For linguistic distance as a contextual variable, we computed the distance from the mother tongue of the L2 learners to Dutch. We did so by extracting branch lengths from the phylogenetic trees of Gray & Atkinson (2003) and ASJP (Wichmann, Müller, et al., 2010), using the APE package (E. Paradis, Claude, & Strimmer, 2004) in *R* (R Core Team, 2013) and dedicated ASJP software (see below). We used the phylogenetic consensus tree of Gray & Atkinson where branch lengths are proportional to substitutions. These branch lengths (technically patristic distances) are based on expert accounts of character substitutions in 200 item word lists, following the comparative method. In the case of language evolution, a character substitution refers to the inferred changes in cognacy status of a word in the Swadesh list. These substitutions, together with an expert-based tree topology, are used by a computational model to infer substitution rates. The computational model of G&A uses relatively recent Bayesian phylogenetic inference methods. ASJP uses the widely used neighbour joining distance-based phylogenetic algorithm.

We used these two phylogenetic trees (ASJP and G&A) in which the length of a branch indicates the amount of evolutionary change between two nodes. A node can either be a leaf of the tree, which is a language as it currently is, or a shared common ancestor between two leaves. The amount of evolutionary change can be considered as a product of time between two nodes and the speed of evolutionary change between those nodes. The sum of branch lengths joining one language to the other (via the most recent common ancestor) represents the amount of evolutionary change between two languages.

We applied software developed by Holman (2010, 2011) and Huff (2010) to the latest version of the ASJP Database (version 13, Wichmann, Müller, et al., 2011) in order to compute ASJP branch lengths. ASJP measures were extracted for all 35 languages. G&A measures were extracted for 30 languages because they were not available for Kurdish, Bosnian, Pashto, Urdu, and Norwegian. The missing scores were imputed using expectation maximization predicted from ASJP measures. Imputing the missing G&A distances had hardly any influence on their mutual dependency.

Figure 1 shows a scatter plot with both linguistic distance measures on the axes. It shows that differences between distances from Romance, Slavic, or Baltic languages to Dutch are fairly small; hence the graph contains a part that zooms in there. The correlation of ASJP with G&A was .90, see Figure 1. In terms of phylogenetic differences between both linguistic distance measures, we see that in G&A, the distance between Germanic and non-Germanic languages is 83% of the average distance from non-Germanic languages to Dutch, whereas in ASJP it is 33%. In G&A, the distance from Germanic languages to Dutch is 16% of the average distance from non-Germanic languages to Dutch, whereas in ASJP, it is 67%. In other words, the distance from Germanic languages to Dutch relative to the distance from all other IE languages to Dutch is 2.5 times higher in ASJP than in G&A. Also, we see that the distance between Germanic and other IE languages relative to distances from other IE languages to Dutch is more than 4 times higher in G&A than in ASJP. Although the correlation between the two measures is high, the underlying deviations from the means differ for relative distances between individual languages as well as between genera. As our results show below, the differences between ASJP and G&A have consequences for the performance of both measures in our regression analyses.

The second contextual level is country of birth. We extracted educational difference measures from the World Bank database. We used gross secondary school enrolment (available for all selected countries) as predictor of educational quality. This variable measures the ratio of total enrolment into secondary education. Secondary education is part of the basic education program that begins with primary education. It offers subject and skill-oriented instruction from specialized teachers. Where available, data from 2006 was used. When 2006 data was not available, earlier data was used. For the former Soviet Union and Yugoslavia, estimates of the current countries were used based on each learner's mother tongue. Schooling quality correlated -.17 with G&A and -.21 with ASJP.

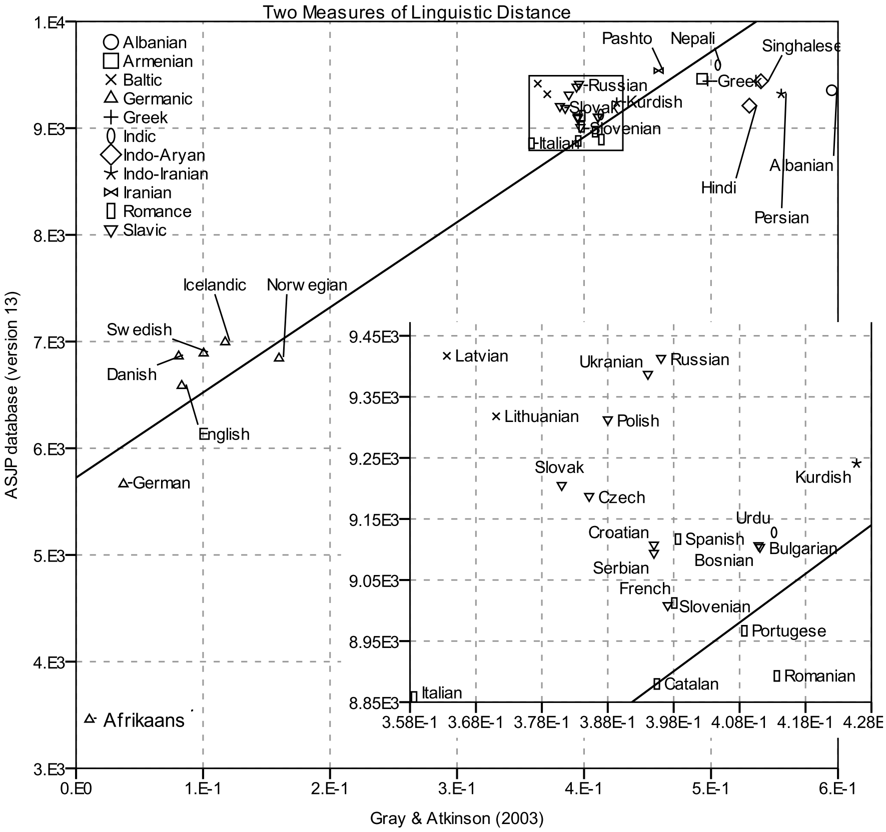


Figure 1. Scatterplot (with a linear regression line) of the two measures of linguistic distance from 34 Indo-European languages to Dutch. Subfamilies are distinguished by using different symbols.

Individual Characteristics

With respect to the capacity of the learner, we added gender, years of full-time education, and a binary indicator whether or not the examinee had already mastered an additional language beforehand. Years of full-time education was measured by asking the examinees, prior to taking the exam, to estimate the number of years that they received full-time education. We measured this variable in steps of 5 years (finer granularity was not possible). The mean years of full-time education was about 13 years. For a more detailed discussion of the variables added, see Van der Slik (2010). Besides adding gender and years of full-time education, we added age of arrival and length of residence as measures of exposure.

Description of the Sample

The temporary increases and decreases of specific groups of examinees taking the state exam tend to overlap with historical events, such as the admission of Poland to the European Union and sharpened rules for marriage across EU borders. The five largest language groups represent almost half of the examinees in our dataset (53.3%). The other language group sizes decrease in a logarithmic fashion.

Most of the sample's learner characteristics, such as mean speaking proficiency, were somewhat lower than those in Van der Slik (2010), as we included also non-Western European countries. Average gross enrolment rate and number of countries with a liberal democracy decreased with respect to the larger data set used here.

Analyses

We first constructed a multilevel model with migrants cross-classified by languages and countries with no predictors added. An analysis of the languages included in our study showed that country characteristics do not necessarily overlap with language characteristics and vice versa. Table 1 exemplifies the Southwest Asian situation in which languages and countries are not uniquely mapped.

Table 1. Cross classification of mother tongue by country of birth in Southwest Asian learners of L2 Dutch. Numbers are based on our dataset. Cells with less than 20 examinees were excluded from the analyses (i.e. reset to zero).

	Kurdish	Farsi	Armenian	Pashto
Iraq	738	0	71	0
Iran	91	2063	45	0
Armenia	0	0	109	0
Afghanistan	0	1252	0	274

To measure the effect parameters of the various determinants we had identified, we added fixed effects to the model step by step. First, the learner characteristics were added to the null model as a baseline model. Then, the contextual determinants were added one by one. Improvement in fit was accepted only when an addition of a new predictor resulted in an improvement of fit of at least a chi-square of 3.84 at p = .05 against 1 degree of freedom on the -2 log-likelihood ratio (-2LL). We will call this the *deviance* between the old and new model. Only when the deviance of the newer model was significantly lower than the deviance of the older model, we checked the direction and size of the individual and contextual effects.

Following Hox (2002) and Heck, Thomas, and Tabata (2010), a cross-classified model of variance components between languages, between countries, and within a language and country together, can be modelled as follows:

$$Y_{i(jk)} = X_{i\,(jk)}\beta_{0\,(jk)} + \varepsilon_{\mathrm{i}(jk)}$$

where $Y_{i(jk)}$ is the proficiency of learner *i* within the cross-classification of languages j and countries *k*; $\beta_{0\,(jk)}$ is the intercept (overall mean proficiency) of learners for language *j* in a country *k*; the residual $\varepsilon_{\mathrm{i}(jk)}$ is the deviation of learner *ijk*'s proficiency from the language j in country k mean. The parentheses indicate that classifications are grouped together at the same level. The model assumes equal variance at the learner level, but still allows predictors to cross-level interact with fixed or random effects at the contextual level 2. Furthermore, the model assumes that proficiency varies independently across languages and countries.

The level 2 null model is:

$$\beta_{0\,(jk)} = \gamma_{00} + u_{0j} + \upsilon_{0k}$$

where γ_{00} is the grand mean proficiency of all learners; u_{0j} is the residual error for language *j* (the contribution of language *j* averaged over all countries), and υ_{0k} is the residual error for country *k* (the contribution of country *k* averaged over all languages).

The application of this null model to the speaking proficiency scores results in three variance components, one for each random effect and one for residual variance. The proportion of variance that is due to differences between languages and countries can be estimated with a measure of the dependency between individual learners, called the intra-class correlation. The between language differences can be estimated by:

$$\frac{\sigma^2_{language}}{\sigma^2_{language} + \sigma^2_{country} + \sigma^2_{e}}$$

where the squared sigmas represent the variance components. The measure indicates that 10.6% of the variation in proficiency scores is across languages and 14.2% is across countries. Summing these up (Goldstein, 2011), we observe that 24.8% of the total variance can be attributed to country and language as characteristics of groups of learners. Accordingly, the remaining variance at the individual level was estimated at 75.2% of the total variance in proficiency scores (these percentages are underlined in Table 2).

In the next section, we will try to explain the reported variance between languages (10.6%). For this purpose, we add fixed level 1 and level 2 explanatory variables to the cross-classified design of languages by countries. The null-model coefficient $\beta_{0\,(jk)}$ gradually becomes a vector of fixed part coefficients by the addition of more variables to the variable design matrix $X_{i\,(jk)}$. Adding one predictor results in:

$$X_{i\,(jk)}\beta = \beta_0 + \beta_1 x_{1i(jk)} + u_{0j} + \upsilon_{0k}$$

where $\beta_1\, x_{1i(jk)}$ is the fixed slope defined by a parameter estimate of a predictor variable.

Results

In this section, we specify the characteristics of the cross-classified multilevel models that we constructed with learner data cross-classified across home countries and mother tongues. The results show that learner and contextual determinants explain part of the variance in speaking proficiency levels within and between groups.[3] First, we report measures of fit resulting from the addition of a number of fixed predictors to the null model. Second, we report how level 1 and level 2 fixed predictors interact. Third, we compute predicted scores based on fitted parameters and compute the correlation of observed scores with predicted scores instead of raw linguistic distance measures. Fourth, we compare parameter settings for G&A with parameter settings for ASJP.

Table 2 shows how the linguistic distance measures correlated with observed speaking proficiency scores at the individual level, the language level, and at the cross-classified level of language by country. From this table it can already be inferred that speaking proficiency is strongly related to linguistic distance.

---

[3] We also tested a number of other contextual effects but these were non-significant (ns) and were therefore excluded from the final model. These effects were: writing system (ns), speaker population size (ns), number of learners in the sample with the same country of birth (ns), and whether or not the country had officially been in a continuous state of liberal democracy during the last 20 years (ns). We also tested if scores differed before and after 2005 (ns). After 2005 immigrants were able to fulfil requirements for a residence permit by completing the state exam instead of the usual lower level naturalization course. Before 2005 this was not allowed. Furthermore, we tested the effect of gross domestic product per capita using data from the CIA (2011). Although this effect was significant ($p<.001$), we excluded it from our final analyses in favour of a simpler model.

Table 2. Correlations of lexical distance measures and speaking proficiency, at the cross-classified level of mother tongues and countries of birth (Co x L1), at the country of birth level (Co), at the mother tongue level (L1), or at the individual learner level (In). N gives the number of cases at the level investigated.

	G&A (L1)	ASJP (L1)	Schooling (Co)	N
Speaking (Co x L1)	-.49	-.49	.67	118
Speaking (L1)	-.77	-.66		35
Speaking (Co)			.66	89
Speaking (In)	-.42	-.40	.32	33,066

Note. All correlations were significant at the .01 level or higher (2-tailed)

Estimated Models

Adding individual determinants resulted in a baseline model that explained part of the variance observed in the null model, as can be seen in Table 3. Adding gender, age of arrival, length of residence, years of full-time education, and command of an additional language reduced the unexplained variance observed in the null model by 4.3% at the individual level, 3.1% between language variance, and 9.8% between country variance. The deviance measure indicated that the model for speaking fitted better to the data than the null model (a decrease of 1,435.5 in the deviance score, with five parameters added).

Adding contextual determinants resulted in a model that explained most of the remaining contextual variance observed in the baseline model. With respect to country level characteristics, most variance was explained using the World Bank measures of gross secondary school enrolment. The C Model (Country Model) fitted better to the data than the Baseline Model (the –2 log likelihood ratio decreased with 54.5 points against one degree of freedom).

After addition of language level characteristics to the country model, we observed significant effects for both measures of linguistic distance. Models at this step contain country (C) and mother tongue characteristics (C+T Models). Both C+T models fitted better to the data than the C Model, see Table 3, because -2log-likelihood ratios decreased with 17.6 (ASJP) points and 26.6 (G&A). Both reductions are significant against 1 degree of freedom ($p<0.05\%$). Because the resulting -2loglikelihood is lower for G&A, we conclude that the G&A model fits better to the data (the critical value for a significant difference is 3.84). The percentage of explained between-language variance rose from 25.3% to 63.7% (ASJP) and to 75.1% (G&A). These differences in explained variance indicated that the G&A based model leaves less variance in the data unexplained.[4]

A multilevel model can allow for the effect of a learner characteristic to vary randomly across languages and countries. Because we are interested in establishing a robust analysis of between-language variation, it is informative to assess whether learner characteristics interact with contextual level characteristics. In this case, we derived from our hypothesis that linguistic distance may enhance the negative effects of age of arrival and length of residence. Hence, we allowed these individual characteristics to interact in a fixed way with the contextual effect of linguistic distance.

We also tested robustness by incorporating a fixed interaction effect between schooling quality and education length, as it is likely that a lower education quality lowers the positive effect of a longer education.

With the addition of these three interactions to the model, we observed a strong overall improvement of model fit. The intercept estimates remained largely the same while the deviance from the data

---

[4] Adding G&A language level predictors without country level predictors resulted in R2 measures of 13.8% (country level), 66.9% (language level), and 4.3% (learner level), implying that language characteristics, and not characteristics of countries, actually explain most of the between-language variance. The predicted scores correlated with .82 ($p<.01$) at the language level, and with .48 ($p<.01$) at the individual level. Adding ASJP language level predictors without adding country level predictors resulted in R2 measures of 11.7% (country level), 48.8 (language level), and 4.3% (individual level). The predicted scores correlated with .73 ($p<.01$) at the language level, and with .45 ($p<.01$) at the individual level.

decreased substantially, indicating that the model fitted better to the data with the addition of interaction variables. The models with interaction effects (C+T+I*C/T Models) both fitted significantly better than the C+T Models. The -2loglikelihood decreased with 505.5 (ASJP) and 452.0 (G&A) points against three degrees of freedom. The explained variance only increased marginally between models with and without interaction effects (1.7% for ASJP and .1% for G&A). The interactions of linguistic distance with length of residence and age of arrival were both significant. Adding these interactions to the model shifted the effects of linguistic distance and age of arrival (in the case of G&A) to non-significant. The third interaction between educational quality and years of education was significant in both models. In all, we found that all three learner characteristics significantly interact with contextual characteristics. With respect to education, a longer education generally has less of an effect as educational quality is lower. The interaction might be a kind of effectiveness measure of received education. With respect to age of arrival and length of residence, being older at arrival and residing for a longer period generally influence second language learning more negatively as linguistic distance is greater. These interactions might imply that coping with a greater distance is more difficult at a later age due to decline of cognitive functions, and when being longer but less intensively exposed.

Table 3. Multilevel model parameter estimations for measures of Dutch speaking proficiency (standard errors in parentheses) per mother tongue and country of birth.

	Null Model	Baseline Model	C Model	C+T Model (ASJP)	C+T + I*C/T Model (ASJP)	C+T Model (G&A)	C+T + I * C/T Model (G&A)
Level 1, learner effects (Baseline Model)							
Intercept	522.42^{3}	517.73 ^{3}	489.87 ^{3}	533.96 ^{3}	473.00 ^{3}	509.97 ^{3}	496.78 ^{3}
	(2.65)	(2.84)	(4.07)	(10.17)	(11.12)	(5.07)	(5.60)
Female (H1)		5.73 ^{3}	5.70 ^{3}	5.72 ^{3}	5.85 ^{3}	5.71 ^{3}	5.75 ^{o}
		(.42)	(.42)	(.41)	(.41)	(.41)	(.41)
Age of arrival (H2)		-.71 ^{3}	-.71 ^{3}	-.72 ^{3}	1.84 ^{3}	-.72 ^{3}	.058 ^{3}
		(.024)	(.024)	(.024)	(.12)	(.024)	(.045)
Length of residence (H3)		.34 ^{3}	.34 ^{3}	.34 ^{3}	.62 ^{2}	.34 ^{3}	.45 ^{3}
		(.040)	(.040)	(.040)	(.22)	(.040)	(.076)
Full-time education (H4)		2.03 ^{3}	2.03 ^{3}	2.03 ^{3}	-1.17 ^{o}	2.03 ^{3}	-.60 ^{o}
		(.21)	(.21)	(.21)	(.73)	(.21)	(.74)
Additional language (H5)		6.80 ^{3}	6.83 ^{3}	6.86 ^{3}	7.07 ^{3}	6.86 ^{3}	7.08 ^{3}
		(.51)	(.51)	(.51)	(.51)	(.51)	(.51)
Level 2, country of birth effects (C Model)							
Educational quality (H6)		–	.34 ^{3}	.34 ^{3}	.21 ^{3}	.33 ^{3}	.22 ^{3}
			(.039)	(.038)	(.04)	(.038)	(.044)

Level 2, mother tongue effects (C+T Model)							
Linguistic distance (H10)		–	–	-5.16E-03 ^{3}	3.51E-03 ^{2}	-54.90 ^{3}	14.32 ^{o}
				(1.08E-03)	(1.17E-03)	(8.79)	(9.46)
*Cross-level interaction effects (C+T + I*C/T Model)*							
Age of arrival (H2) * Linguistic distance (H10)				–	-3.15E-04 ^{3}	–	-2.56 ^{3}
					(1.50E-05)		(.13)
Length of residence (H3) * Linguistic distance (H10)				–	-3.24E-05 ^{o}	–	-.30 ^{o}
					(2.60E-05)		(.21)
Full-time education (H4) * Educational quality (H6)				–	3.93E-02 ^{3}	–	3.34E-02^{3}
					(8.03E-03)		(8.08E-03)
Variance components							
Learner	977.51^{3}	935.73^{3}	935.86^{3}	935.96^{3}	921.68^{3}	935.96^{3}	923.21^{3}
	(7.65)	(7.32)	(7.32)	(7.33)	(7.21)	(7.32)	(7.22)
Country of birth	184.33^{3}	166.35^{3}	79.07^{3}	76.14^{3}	72.50^{3}	75.48^{3}	71.72^{3}
	(34.68)	(31.51)	(16.30)	(15.44)	(14.80)	(15.10)	(14.5)
Mother tongue	137.28^{2}	132.98^{2}	102.59^{2}	49.80^{2}	52.33^{2}	34.20^{1}	34.27^{1}
	(48.45)	(46.73)	(33.01)	(19.03)	(19.50)	(13.98)	(13.89)

Measures of fit							
$R^2_{Learner}$	0 (<u>75.2</u>%)	4.3%	4.3%	4.2%	5.7%	4.2%	5.6%
$R^2_{Country}$	0 (<u>14.2</u>%)	9.8%	57.1%	58.7%	60.7%	59.0%	61.1%
$R^2_{Language}$	0 (<u>10.6</u>%)	3.1%	25.3%	63.7%	62.0%	75.1%	75.0%
-2LL	319,076.0	317,640.5	317,586	317,568.4	317,063.9	317,559.4	317,107.4
Correlations of fixed predicted values with proficiency measures							
Mother tongues	0	.32º	.72³	.84³	.84³	.87³	.86³
Learners	0	.24³	.40³	.49³	.50³	.49³	.50³

Legend: Reference categories are Male, Monolingual; effects B; º $p \geq .05$, ¹ $p < .05$; ² $p < .01$; ³ $p < .001$.

Fixed Predicted Scores

The discussion of the variance components suggested that most of the variance across mother tongues could be explained by the fixed effect parameters we fitted to the data. The remaining variance across mother tongues suggests that the model's fixed predicted scores do not completely overlap with observed scores. Here, we inspect this overlap at the level of the mother tongue. The fixed predicted scores can be inferred using the fixed effect parameter estimates. The fixed predicted scores are essentially regression means over the remaining random variance in the model, represented by the variables and their parameters only. These scores can be averaged over languages to inspect predicted differences across mother tongues and assess if they overlap with observed differences.

Mean observed scores and fixed predicted scores are shown in Figure 2 for both ASJP (left) and G&A (right). A linear regression line represents the linearity of the model predictions, which we applied for all parameter estimations. The model predictions show in detail how every single unit deviates from the linear regression line, enabling quantitative comparison between models and predictions. The points deviate from the linear fitted line in comparable ways between both models. For example, speakers of Kurdish seem to score far under their predicted score in both models. Speakers of German performed even better than inferred from their favourable parameter settings.

A closer look at both panels also reveals a number of differences between the two different models. We consider a number of differences between both models, and judge the correctness of their claims according to the distance of the prediction to the linear model and the distance with each other. In Table 4, we show the ten languages on which both models disagree the most (*model difference*). In Figure 2, model difference is represented as the difference in position on the x axis for a language. For example, Albanian has the highest difference on the x axis. The other column in Table 4 shows which of the models is more accurate in terms of observed proficiency scores (*difference in fit*). For example, G&A found a better fit for Albanian of 2.77 points because the ASJP based estimation was 5.94 too high, whereas the G&A bases estimation was 3.17 too low. For Afrikaans, the observed score did not provide much evidence for either the one or the other prediction as both models deviate about equally in different directions from the observed score. Summing up all the differences in fit, the G&A model fitted 14.36 points better than the ASJP predictions (average per language of .41). The differences spread about equally across linguistic subgroups.

Table 4. The 10 most different predictions between the ASJP and G&A models. Positive values indicate difference in fit in favour of G&A, negative values indicate difference in fit in favour of ASJP.

Language	Difference in fit	Model difference
Albanian	2.77	9.11
Afrikaans	-0.81	7.53
Persian	5.12	6.86
Hindi	-6.09	6.09
Danish	4.18	5.92
Singhalese	5.64	5.64
Swedish	5.15	5.15
Icelandic	-4.71	4.71
English	-4.57	4.57
Latvian	3.82	3.82

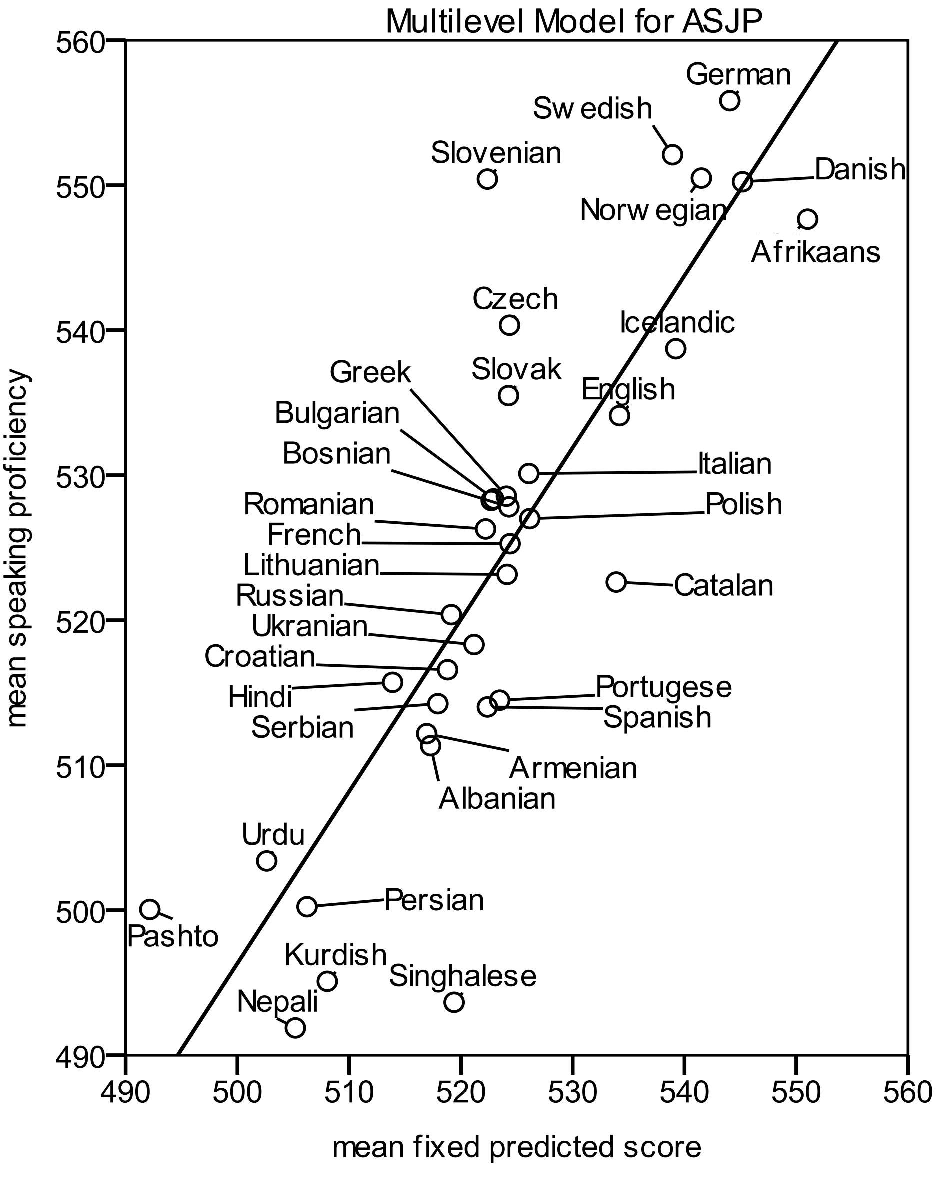
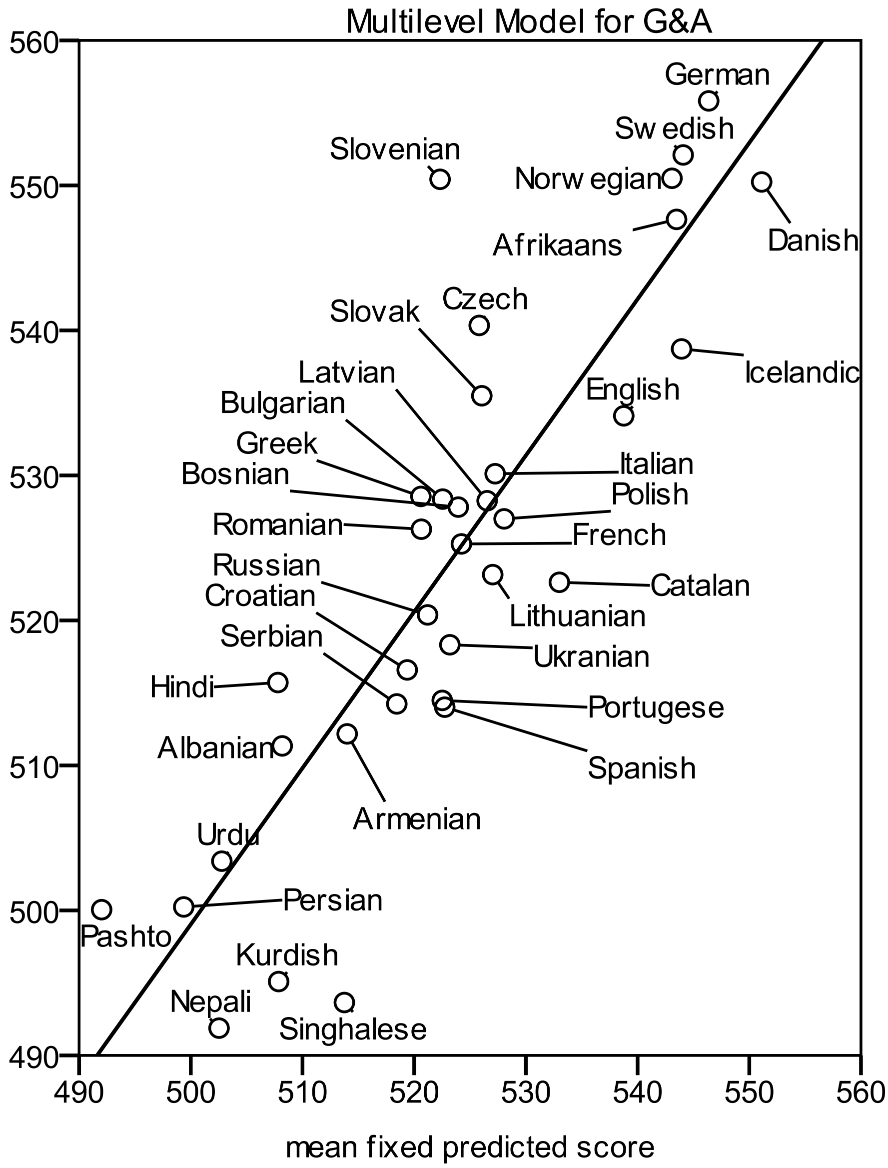


Figure 2. Language level fixed effect part estimates of the multilevel interaction model for speaking proficiency. The first panel shows estimates for ASJP measurements;



the second panel shows estimates for G&A measurements. Deviations from the fitted line represent either a higher observed speaking proficiency than predicted speaking proficiency or vice versa.

Model Parameter Comparison

We can now take a closer look at both models and inspect their specific parameter settings as depicted in Table 3. The predicted values have shown us that the G&A model generally provides better estimates than the ASJP model. An inspection of the model parameters can provide additional information on the nature of these estimates. To do so, we look at estimated effect sizes B, while keeping in mind that the ASJP model explains about 11.5% less variance across mother tongues than the G&A model (75.1 for G&A - 63.7 for ASJP).

G&A and ASJP models behaved differently in terms of their level 1 intercept estimations. Adding interactions reduces this difference to some extent. However, estimated intercept size is not very meaningful for comparing models. In general, individual learner effect estimations were already well established in the baseline model and did not change much by the addition of level 2 predictors to the baseline model. As in Van der Slik, 2010, we found an advantage for being female over being male, arriving younger, having resided longer, having full-time education for a longer period, and having command over an additional language besides Dutch and the mother tongue. Adding these predictors resulted in a model that can account for confounding variables.

With respect to the level of the country of birth, no large differences were found between the models tested. The interaction of full-time education with educational quality was constant between the ASJP and G&A models. In both models, the individual effect of full-time education became non-significant with the addition of the interaction effect. The proposed meaning of this interaction as a measure of educational effectiveness seems to account for the effect of years of full-time education.

With respect to the level of the mother tongue, linguistic distance brought the unexplained between language variance to a minimum. The different effect sizes of ASJP and G&A are difficult to compare because both follow different scales. However, given that all the other variables are identical across models, it is allowed to use the difference in variance components and therefore the percentages of explained variance to indicate differences in fit to the data between the two linguistic distance measures. As the percentage of explained variance is higher for G&A's measure of linguistic distance, we claim that this measure behaves best in terms of fit to the data.

Altogether, the interaction models incorporated 13 parameters, of which 11 were fixed and 2 were random. Educational quality explained most of the variance across countries of birth. The linguistic distance measures explained most of the variance across mother tongues.

Discussion and Conclusion

We investigated the effects of two linguistic distance measures on the variation in speaking proficiency scores across 30,066 learners, having 35 different mother tongues, originating from 89 different countries (resulting in 119 language by country subgroups). We fitted a range of fixed learner level, country level, and language level effects with either the G&A or ASJP linguistic distance measure to the observed scores. Thereafter, we compared estimated model predictions of mean scores by mother tongue against observed means by mother tongue. In this section, we discuss how the multilevel model settings relate to learning effects of linguistic distances in general. We look in more detail at the levels that we analysed, and more specifically at learning difficulty and linguistic distance.

We started the analysis by distinguishing variance components on three levels in the null model. *Intra-unit correlations* indicated that 10.6% of the variation in scores varied across mother tongues, 14.2% varied across countries, and 75.2% varied at the individual level. A cross-classification analysis on the level of country and mother tongue allowed us to distinguish these two effects and to separate the impact of language on the basis of linguistic distances. The effect of distance from the mother tongue to Dutch was consistently found in models with this kind of structural hierarchy.

The final multilevel models incorporated significant individual learner effects (gender and additional language) and cross-level interactions of age of arrival with linguistic distance, length of residence with linguistic distance, and years of full-time education with educational quality. The small decline in explained variance when adding interaction effects (63.7 to 62.0 for ASJP, 75.1 to 75.0 for G&A) shows that the interaction slopes explain slightly less variance of the overall between-language variance. In general, the negative interaction effects indicate that being younger and having resided for a shorter period in the host country together results in higher proficiency scores, while being older and having resided in the Netherlands longer results in lower proficiency scores. One explanation of these interactions is that the estimation procedure found a dependence of relatively small distance with relatively young age in the special case of German learners. Incentives, which may or may not be an important category of predictors, might be relatively high in German learners because of a substantial degree of university attendance that is present in this group. However, the effect of linguistic distance keeps its robust and pivotal place in explaining variance between mother tongues across models regardless of the interaction effects that we fitted to the data.

A significant percentage of variance in speaking proficiency scores could be ascribed to differences in mother tongues (10.6% of the total variance across learners). The lowest observed mean for a mother tongue was observed for Nepali (491.9 points) and the highest one for German (555.8 points), resulting in a difference of 63.9 points on the scoring scale (see Figure 2). Because linguistic distance explained most of this variance component between languages, we conclude that linguistic distance nicely predicts general difficulty of learning Dutch as a second language. More specifically, we conclude that learning difficulty gradually increases with a higher linguistic distance. We expect that a deeper understanding of the differences between languages requires a more detailed model of linguistic effects (e.g., by including other linguistic distance measures), and a more complete model of learner effects (e.g., by including linguistic distances of additional L2s, acquired before arriving in the Netherlands). Given that addition of cross-level interactions to the model explained more learner level variance than language level variance, we expect that further modeling of cross-level effects will enhance the model's performance at the learner level, leaving the explained variance between languages more or less intact. The cross-level interaction effect between age of arrival and linguistic distance did not add to the degree of explained variance between languages while it explained a substantial amount of variance between learners within languages.

We have seen that both an automatic and an expert-based linguistic distance measure are appropriate instruments to explain most of the empirically observed between-language variation across learners. The predicted scores were in favour of the G&A distances (better average fit of .41 points per language and a difference in explained variance of more than 10%) than the ASJP distances. Given the differences between the two measures described earlier, this finding suggests that distances from Germanic to Dutch are relatively small and distances from Germanic to other Indo-European languages are relatively great if they are used for explaining the linguistic distance effect in SLA. However, in terms of proficiency scores, both the mean observed and mean predicted proficiency scores develop more gradually than both measures of linguistic distance do. We hope to investigate the role of linguistic distance further by turning the model around. Can we predict the optimized distances from a reversed model? Such a reversed measure may inform us whether empirically determined linguistic distances are distributed differently from phylogenetically determined linguistic distances. Interesting testing cases are, in many respects, the non-Indo-European languages in our database.

We conclude that linguistic distance measures are impressive predictors for explaining average differences in L2 speaking proficiency scores between Indo-European mother tongues (63.7% for ASJP and 75.1% for G&A). This outcome is remarkably robust against more complex models. The correlation between the mean scores of learners of Dutch as an L2 with the distance from their mother tongue to Dutch starts at .66 for ASJP and .77 for G&A. No other variables were included in the computation of these correlations, but these raw correlations support the idea that linguistic distance and L2 learning are related. Incorporating other effects, both on the country level and the learner level, raised the correlations substantially: to .84 for AJSP and .87 for G&A (see Table 3). These high correlations provide convincing evidence that linguistic distance is an important factor in SLA. Addition of cross-level interaction effects led to an improvement of fit while the effect size and relative ordering of linguistic differences remained consistent.

Chapter 3

Learning Complex Features: A Morphological Account of L2 Learnability

Acknowledgments
We are grateful to the secretary of the State Examination of Dutch as a Second Language for providing the data used in our study. We thank Kate Bellamy, Michael Dunn, Michele Gubian, Gary Lupyan, Theo Bongaerts, and three anonymous reviewers for comments on an earlier version of this manuscript and editorial help.

This chapter was published as:
Schepens, J., Van der Slik, F., & Van Hout, R. (2013). Learning complex features: A morphological account of L2 learnability. *Language Dynamics and Change, 3(2)*. Special issue, "Phylogeny and Beyond: Quantitative Diachronic Approaches to Language Diversity", edited by Michael Dunn.

Abstract

Certain first languages (L1) seem to impede the acquisition of a specific L2 more than other L1s do. This study investigates to what extent different L1s have an impact on the proficiency levels attained in L2 Dutch (Dutch L2 learnability). Our hypothesis is that the varying effects across the L1s are explainable by morphological similarity patterns between the L1s and L2 Dutch. Correlational analyses on typologically defined morphological differences between 49 L1s and L2 Dutch show that L2 learnability co-varies systematically with similarities in morphological features. We investigate a set of 28 morphological features, looking both at individual features and the total set of features. We then divide the differences in features into a class of increasing and a class of decreasing morphological *complexity*. It turns out that observed Dutch L2 proficiency correlates more strongly with features based on increasing morphological complexity ($r = -.67$, $p < .0001$) than with features based on decreasing morphological complexity ($r = -.45$, $p < .005$). Degree of similarity matters ($r = -.77$, $p < .0001$), but increasing complexity seems to be the decisive property in establishing L2 learnability. Our findings may offer a better understanding of L2 learnability and of the different proficiency levels of L2 speakers. L2 learnability and L2 proficiency co-vary in terms of the morphological make-up of the mother tongue and the second language to be learned.

Morphological complexity, WALS, adult language learning, L2 learnability, speaking proficiency

Learning Complex Features: A Morphological Account of L2 Learnability

Introduction

Children seem to learn languages easily, in a natural way, unlike adults, who often struggle when learning to understand a second language and express themselves in it. Their struggle can often be noticed in their use of L2 morphology, as inflected forms are often missing or incorrect (for L2 Dutch, see (Oldenkamp, 2013). Previous research on L2 learning impediments has taken different perspectives on L1-L2 linguistic differences, for example by means of (1) contrastive analysis (Lado, 1957; Odlin, 1989; Towell & Hawkins, 1994; Weinreich, 1963), (2) linguistic distance (Chiswick & Miller, 2005; Van der Slik, 2010), and (3) morphological complexity (Dahl, 2004; Lupyan & Dale, 2010; McWhorter, 2007; Nettle, 2012).

The notion of morphological complexity is relevant for explaining patterns of variation in the morphological make-up of languages. Language contact has a direct impact on morphological complexity, in particular in combination with mechanisms of adult language learning. Correlational evidence obtained from typological data (Lupyan & Dale, 2010; Nettle, 2012) indicates a decrease in morphological complexity of languages when the number of L2 learners increases. These studies confirm on a larger scale what is observed in smaller scale acquisition studies (Ionin & Wexler, 2002; Lardiere, 1998): adult learners have persistent problems in L2 acquisition, especially in acquiring L2 morphosyntax.

If complexity is so essential, it is tempting to conclude that some languages are easier to learn for adults than others are. Trudgill (1983, 2011) points to (Dauenhauer & Dauenhauer, 1998), who investigated reversing language shift in Tlingit, Haida, and Tsimshian. They conclude “the languages of Southeast Alaska are intrinsically more difficult to learn than Maori or Hawaiian because of their more complex grammars and phonologies.” Variation between languages in their morphological make-up and complexity has a strong influence on how these languages are transmitted in language contact scenarios (Andersen, 1988; Braunmüller, 1990; Dahl, 2004; Kusters, 2003). The consequences are, as Trudgill argues, that the 'easier' languages are highly analytical (less complex morphology, more lexical means), often because they have experienced more contact. Language complexity is linked to adult L2 learning difficulty, although the precise mechanisms involved are far from clear. As Trudgill (2011: 41) notes, "Dahl (2004: 39) prefers to suppose that complexity and L2 difficulty are not actually identical but simply 'related.'"

The main aim of the present study is to investigate whether data regarding adult L2 learning, in particular L2 Dutch, reveal effects of morphological distance (differences) and complexity. We have shown earlier that the lexical distance between L2 Dutch and the L1s of L2 learners is systematically correlated with L2 Dutch proficiency (Schepens, Van der Slik, & Van Hout, 2013b). Secondly, we want to investigate the additional value of morphological L1-L2 distance measures compared to the impact of lexical L1-L2 distance we found earlier. We hypothesize that differences in morphological make-up in general and differences in morphological complexity in particular account for the L2 learnability of Dutch. More specifically, we expect that L2 learnability is lower when the L1 is morphologically less complex as compared to the L2.

The notion of L2 learnability may help to shed more light on the likelihood of L1-dependent biases in learning L2 linguistic features. Typologically relevant linguistic features for many languages can be found in the online World Atlas of Language Structures (henceforth WALS) database (Dryer & Haspelmath, 2011). Lupyan and Dale (2010) used the WALS data to define a set of 29 morphological features on which they based their correlational study on language structure and population sizes. They ordered the variants of those features (i.e., the feature values) on a complexity scale. We employ the morphological set of features they extracted from the WALS database, also making use of the complexity scales they defined. We systematically compare the

Dutch variants of the morphological features with the variants in the L1s. For every feature in the L1 involved, we check if its variant is morphologically identical, more complex, or less complex as compared to Dutch.

Thus far, there are no large-scale correlational studies of L2 learnability bias in adult L2 learning that encompass the structure of L1s. The correlational study of Lupyan and Dale (2010) implicitly assumes that all L1s are equally responsible for effects of population size on morphological complexity. A strong point of the present study is that we relate L2 proficiency scores to the structural features of the L1s of learners of L2 Dutch. The concept of L1-dependent L2 learnability can thus shed more light on the likelihood of L1-dependent biases in the L2 learnability of linguistic features.

To determine L2 proficiency levels in Dutch for learners who speak a typologically wide variety of L1s, we use speaking proficiency scores of Dutch as an L2 for speakers of 73 different L1s. The database allows for evaluation of morphological distance and complexity by means of a statistical analysis of more than 50,000 L2 proficiency scores.

In the following section, we define morphological complexity and provide an overview of current evidence for the relationship between morphological complexity and adult language learning. In the methods and results sections, we describe the development and testing of the impact of morphological distance and morphological complexity. We test the benefit of morphological distance in relation to lexical distance between Dutch and the L1s involved. In the final section, we discuss our findings and present directions for further study.

Background

When a set of morphological features is available for a set of languages, distances in terms of differences can be counted in a straightforward way by establishing whether the languages in question have the same feature value or not. Making comparisons in terms of morphological complexity is more difficult, however.

Morphological complexity can be defined as the extent to which a language makes use of modifications of words (Nettle, 2012). This definition is in accordance with the notion of structural complexity of linguistic expressions (Dahl, 2004), and fits information theory in terms of compressibility (Juola, 1998; Lupyan & Dale, 2010). It is also in accordance with the notion of complexity in terms of L2 acquisition difficulty (Kortmann & Szmrecsanyi, 2012; Kusters, 2003, p. 6). Complexity reflects the investment needed for an adult L2 learner to acquire another language. It quantifies languages with more inflectional morphology as more complex relative to more isolating languages, based on the assumption that morphology is harder to acquire in an L2 than in an L1.

WALS provides data in terms of feature values across languages with varying degrees of inflectional morphology. Consider person/number marking on the verb, for example. Many languages mark person and number of the subject on the verb; however, in languages of Southeast Asia this is quite uncommon, as can be seen in Figure 1. WALS contains at least 29 morphological features whose values range from less complex lexical variants to more complex inflectional devices (Lupyan & Dale, 2010). An overall degree of morphological complexity can be obtained by pairwise comparisons of the morphological complexity of feature values. Using the lexical-inflectional rank orders given in Table 1 of Lupyan and Dale (2010) as scales, languages can be compared and evaluated in terms of their morphological complexity.

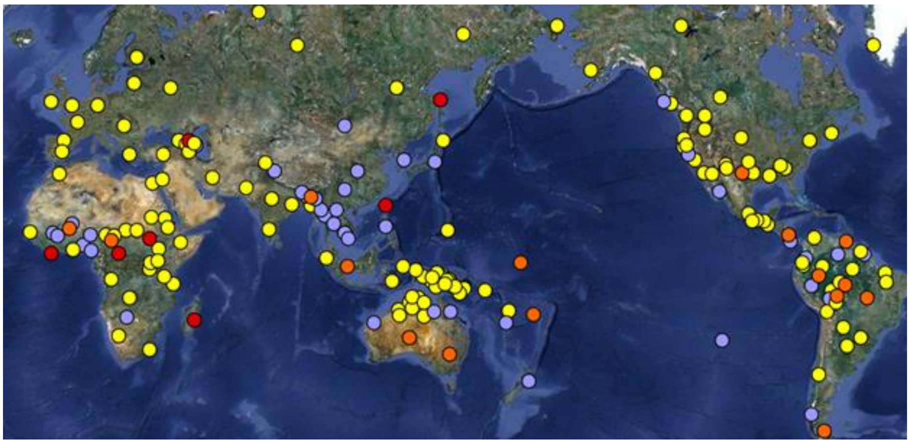


Figure 1: Verbal person marking (100): neutral (violet and red) versus non-neutral alignment (yellow and orange). Verbal subject marking for person and number (feature 29): none (violet and orange) versus other than none (yellow and red) (Dryer and Haspelmath, 2011)

This approach to morphological complexity challenges the traditional view that structural complexity is distributed uniformly across languages (Hockett, 1958, p. 180). It allows for cultural-evolutionary mechanisms that affect the development of complexity (Sampson, Gil, & Trudgill, 2009). There is, in fact, recent evidence for the existence of cultural-evolutionary mechanisms in language structure (Evans & Levinson, 2009). For example, differences in language structure may be due to differences in genetic bias (Dediu & Ladd, 2007; Hunley et al., 2008) and population size (Wichmann & Holman, 2009; Wichmann, Stauffer, Schulze, & Holman, 2008). Thanks to, in all likelihood, the better availability of typological databases such as WALS, researchers are beginning to quantify structures cross-linguistically on a large scale.

Table 1 highlights the distinctions between morphologically less and more complex dimensions of language according to the linguistic niche hypothesis of Lupyan and Dale (2010). These authors hypothesize that the differences in social structure between esoteric and exoteric niches affect language structure. Languages with a relatively high number of L2 learners, as found in the exoteric niche, are more likely to use lexical means of expression. In contrast, languages spoken in the esoteric niche are supposedly more complex morphologically, as they adapt to an L1-facilitative structure.

Table 1. Dimensions in which morphologically more and less complex languages are assumed to differ.

Dimension	Morphologically less complex	Morphologically more complex
Restrictedness	Ambiguous	Overspecified
Linguistic Strategy	Lexical / word order	Inflectional / conjugational
Learning Mechanism	Selection (facilitates L2)	Redundancy (facilitates L1)
Linguistic Type	Isolating	Synthetic
Cultural Type	Exoteric	Esoteric
Population	High, many adult learners	Low, many child learners

This observed negative relationship between population size and the degree of morphological complexity is in accordance with research from multiple disciplines. Studies in historical linguistics show that within many language families, morphological inflection has been lost because of changes in community structure (Kortmann & Szmrecsanyi, 2012; Kusters, 2003; McWhorter, 2002, 2007, 2011; Miestamo, Sinnemäki, & Karlsson, 2008; Trudgill, 2001, 2002, 2011). Breaking down population size into specific L1/L2 community size estimates confirms the importance of the number of L2 learners compared to the whole population size (Bentz & Winter, 2013). Psycholinguistic studies and studies in language acquisition have come up with abundant evidence of learning differences between children and adults (Blom, Polisšenská, & Weerman, 2006; Flege, Yeni-Komshian, & Liu, 1999; Johnson & Newport, 1989; McDonald, 2000; Prévost &

White, 2000). In addition, artificial language learning studies have uncovered a weaker bias for regularization in adult language learners as compared to child language learners (Smith & Wonnacott, 2010).

Mandarin Chinese L2 Dutch further illustrates L2 learnability differences with respect to the expression of verbal inflection. Since no verbal inflection exists in Mandarin, one would expect these learners to prefer short verb forms corresponding to the stem of a verb. Oldenkamp (2013: 53) showed that Mandarin Chinese L2 learners of Dutch use verbal inflections less than Moroccan Arabic L2 learners of Dutch (whose native language does have verbal inflection). Hence, the realization of inflection in the L2 may depend on the degree of inflection in the L1.

In Chapter 2, we showed that state exam data can be used successfully to compare how well lexical measures of linguistic distance explain differences in proficiency in L2 Dutch. Two different lexical measures of linguistic distance between the L1 and L2 were tested for their explanatory value of L1 variance in L2 proficiency scores (Gray & Atkinson, 2003; Holman et al., 2008). It was concluded that the effect of the L1 for learning L2 Dutch is a distance effect, as the linguistic distance between the L1 and L2 explained differences in L2 proficiency to a large extent (75.1%). This success raises the question whether morphology can explain the L1 variance in L2 proficiency levels even better. Does it have additional value?

Our first hypothesis is based on the observation that differences in morphological distance and complexity across L1s exist, and the premise that the more inflectional morphology an adult language learner needs to acquire, the lower L2 learnability is. Morphological distance is a result of either more or less morphology between an L1 and an L2. As a baseline, we expect that a higher distance between the L1 and the L2 relates to lower L2 learnability, but that such a distance effect can be explained better in terms of complexity.

More specifically, we expect that the impact of morphology on L2 learnability is consistently present across families despite family-specific biases in the morphological make-up of languages. Recent studies show how some features are more stable than others (Dediu & Levinson, 2012) and how feature distributions depend on lineage-specific trends (Dunn et al., 2011); see for an overview Wichmann (In press). We therefore expect the impact of morphological differences to vary depending on the lineage in which the features evolved. Although we assume that an L2 learnability bias itself is not lineage-specific, a family bias is likely to affect its impact and could potentially conceal effects of morphological differences on L2 learnability.

Furthermore, measures of morphological distance or complexity may explain why a strong effect of lexical distance on L2 learnability can be observed across Indo-European languages. We hypothesize that morphological differences explain differences in L2 proficiency scores better than current measures of lexical distance.

Methods and Data

Proficiency Scores of L2 Dutch

A unique database is available in the Netherlands, consisting of L2 proficiency scores for the state exam *Dutch as a Second Language* for more than 50,000 participants. The exams are administered by the official Board of Examinations in the Netherlands, and developed by a large test battery constructor (Central Institute for Test Development; Cito) and the independent Bureau of Intercultural Evaluation. The exam is tailored to higher education; passing it is a requirement for individuals wanting to obtain admission to certain Dutch educational programs. The full exam consists of speaking, writing, listening, and reading tasks, for which proficiency scores are available for most participants. The speaking part of the exam comprises 14 tasks that are similar to one another, in which participants are required to provide information, give instructions, etc., and has to be completed in 30 minutes. Two independent examiners evaluate the spoken language on both content and correctness according to a formal protocol. The pass level is upper-intermediate, comparable to the B2 level of the Common

European Framework of Reference for Languages: Learning, Teaching, Assessment (Council of Europe, 2001).

Using the results of speaking exams for L1s for which at least 20 L2 proficiency scores were available, it is possible to compare 73 languages (L1s) with Dutch (L2). Following WALS (Dryer and Haspelmath, 2011), the 73 L1s come from 35 different genera which belong to 14 language families. Of these 73 languages, 39 are Indo-European and 34 are non-Indo-European. In the latter group, we have eight Niger-Congo languages, six Afro-Asiatic, four Austronesian, three Altaic, three Uralic, two Dravidian, and two Creole languages (Haitian and Papiamentu), as well as one Kartvelian (Georgian), one Austro-Asiatic (Vietnamese), one Sino-Tibetan (Chinese), and one Tai-Kadai (Thai) language, and, finally, Japanese and Korean.

The L2 proficiency scores were annotated with control variables taken from questionnaire information on gender, educational level, length of residence, age at arrival in the Netherlands, and additional language background(s). Enrollment levels in higher education in the country of origin (UNESCO, 2011) were included as well. We calculated *adjusted proficiency* scores for each L1 language. Adjusted proficiency is the by-L1 adjustment (BLUP) as taken from a multilevel model with the control variables as fixed effects and random effects for the L1 (mother tongue), L2 (additional language acquired before learning L2 Dutch), L1-L2 combinations, and countries (Schepens, Van der Slik, & Van Hout, submitted). The adjusted proficiency measures were extracted with the function *ranef* from the R (R Core Team, 2013) lme4 package (D. Bates, Maechler, & Bolker, 2011). The distribution of the adjusted proficiency scores for the 73 L1s is visualized in Figure 2, where zero indicates the average adjusted proficiency score across L1s. Figures 3 and 4 provide an overview of the L1-specific L2 Dutch adjusted speaking proficiency scores for non-Indo-European and Indo-European L1s, respectively. The proficiency scores are generally higher for Indo-European languages (Figure 4); some exceptions are the Uralic languages, which score higher than many Indo-European languages, and Singhalese, an Indo-European language, whose score is among the lowest overall. In the present study, the adjusted proficiency scores are used as the dependent variable.

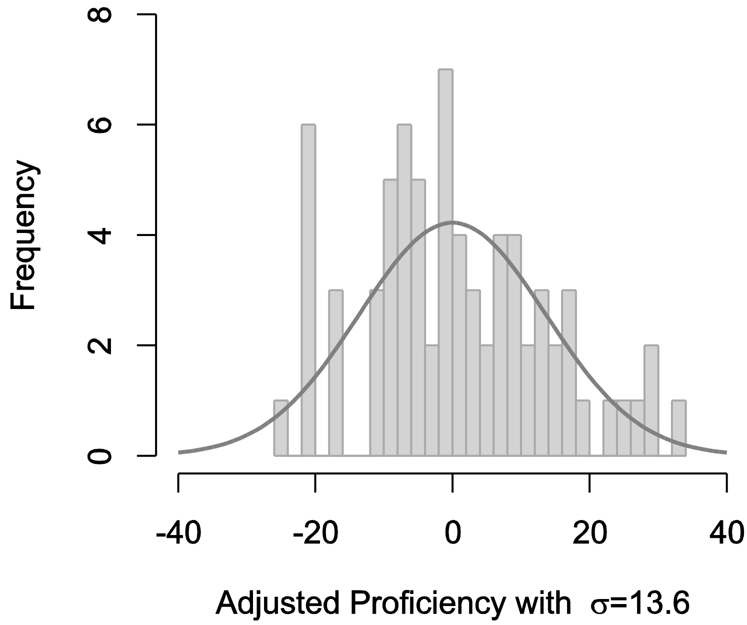


Figure 2. The distribution of adjusted proficiencies exhibits positive skew.

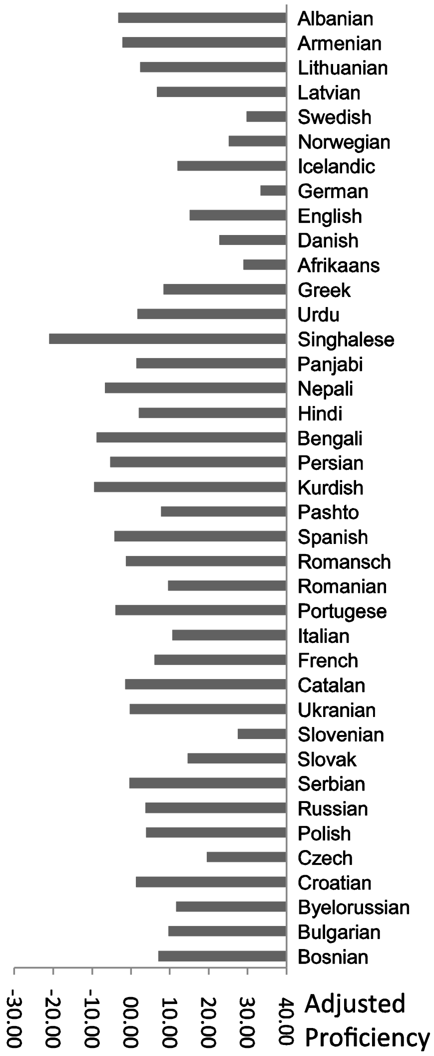


Figure 3. The distribution of adjusted proficiency among the 33 non-Indo-European languages from the 13 non-Indo-European families included in our study. Adjusted proficiency is displayed here relative to that of Estonian, which displayed the highest level of adjusted proficiency (17.55) among non-Indo-European languages.

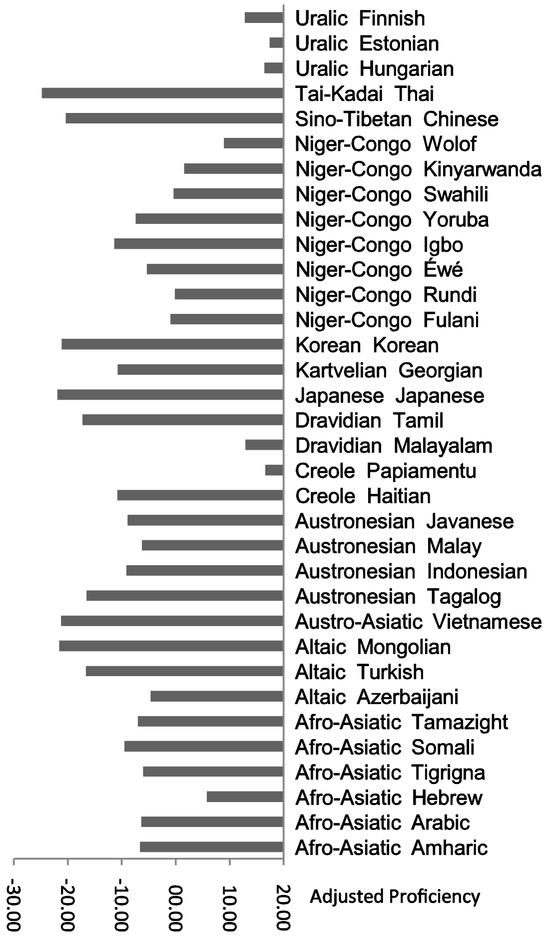


Figure 4. The distribution of adjusted proficiency among the 39 Indo-European languages included in our study. Adjusted proficiency is displayed relative to that of German, which displayed the highest level of adjusted proficiency (33.54) among Indo-European languages.

Morphological Feature Values

Typological features are structural properties of language that represent dimensions of cross-linguistic diversity (Dryer & Haspelmath, 2011). A subset of 29 morphology-specific feature values was extracted from WALS by Lupyan and Dale (2010; note that feature number 26 in their ordering involves two WALS features). These 29 features cover a broad range of morphological dimensions (e.g., agreement, verb inflection, articles) and feature markings (e.g., no plurality vs. obligatory plurality). For our study, we first retrieved all the available feature values from WALS for all the 74 languages in our set (73 L1s plus Dutch). This resulted in a set of 1123 values, excluding all the missing feature values. We filled in six missing feature values for Dutch on the basis of the information provided by the ANS (Algemene Nederlandse Spraakkunst = "Dutch General Syntax"; Haeserijn, Romijn, Geerts, de Rooij, & Van den Toorn, 1997).

All feature values of all the languages included were transformed into three measures in comparing Dutch and the 73 L1s: similarity, increasing complexity, and decreasing complexity. Similarity is 1 for an identical value of a feature and 0 for any other value. Increasing complexity is based on the observed patterns reported in Table 1 of Lupyan and Dale (2010). The measure distinguishes between languages that are less complex than Dutch for a specific feature versus languages that are equally or more complex than Dutch is. The score of 1 indicates that a value in a specific L1 is either equal to Dutch or higher in the complexity ordering, the value of 0 indicates that a value of a specific L1 is lower in the complexity ordering than Dutch. The third transformation defines decreasing complexity from the perspective of the L1s, distinguishing between an equal or lower L1 complexity (coded as 1) versus a higher L1 level of complexity (coded as 0). In all three measures, the 1 is used to indicate equal and the 0 to indicate a difference. It is possible to compare the correlations between adjusted proficiency and each of the three measures in order to test which measure best explains variance in proficiency scores.

Lupyan and Dale (2010) report one feature pattern that seems reversely related to a complexity ordering, namely WALS feature no. 34: Coding/Occurrence of Plurality. Lupyan and Dale's analysis indicates that obligatory plurality marking is more likely for languages in the exoteric niche. This runs counter to the fact that, according to the linguistic niche hypothesis, exoteric languages are generally more likely to use lexical strategies. In contrast, optional plurality marking (using either a word, affix, or clitic) or no plurality marking is more likely for languages in the esoteric niche. This seems to be a contradiction, considering the linguistic properties of rank ordering of the other features. If languages with obligatory plurality, like Dutch, are considered more complex than languages with no or optional plurality marking – contrary to the findings of Lupyan and Dale (2010) – plurality marking correlates strongly with proficiency scores (.651, $p < .0001$, N = 34). It is thus debatable whether obligatory marking is of high or low complexity, and we remove this feature from the set of features analyzed here, resulting in a set of 28 features.

In 12 out of the 28 features considered in the present study, no other language is less complex than Dutch. Examples of languages that, for (almost) every feature value present in Dutch, have a feature value that is either equally or more complex are Hungarian (13 equally or more complex out of 15 observed values), German (12 out of 16), and Georgian (10 out of 13). On the other hand, languages in which feature values of lower complexity, as compared to Dutch, are predominant include Tagalog (9 less complex values out of 14), Vietnamese (12 out of 15), and Chinese (10 out of 12). We first report one-by-one comparisons of feature patterns to proficiency and then evaluate overall measures of linguistic distance based on a combination of these patterns.

Data Analysis

Beyond straightforward morphological similarity between L1 and L2, L2 learnability involves learning an increasing (higher) or decreasing (lower) level of morphological complexity, depending on the morphological features of the L1 and L2. We test whether increasing and decreasing complexity produce significant differences in L2 learnability and whether such a distinction is superior to a similarity-based morphological analysis of L2 learnability.

For evaluating the patterns of individual features, we compute feature-specific point-biserial correlation coefficients between adjusted proficiency scores and the three types of binary feature values (similarity, increasing complexity, decreasing complexity). We compute distances for all three types using a *sum of weighted features.* The weights are the correlations of each morphologically different feature that is more complex in Dutch. We divide these sums by the number of features for which information was available in WALS.

The distance scores were added to the original dataset. We fit linear mixed effects regression models (LMEM) to the adjusted proficiency scores using the *lme4* package in R. We model adjusted proficiency as a function of each of the three distance measures separately in three specific models with one fixed effect each, including a random effect for language family, as well as random slopes quantifying the by-family variance in proficiency. LMEMs can be used for modeling nested dependencies in random variance at the family level (Atkinson, 2011). Separate regressions for each family suffer from data sparseness and are likely to reveal family-specific idiosyncrasies (Jaeger, Graff, Croft, & Pontillo, 2011; Levy & Daumé, 2011). By-family variance is the result of family-specific bias causing languages within a family to be more similar to each other. LMEMs control for such bias by fitting random intercepts and slopes. The random intercepts reflect, by assumption, the normally distributed family-specific intercepts, which capture systematic deviations in proficiency from the average family. The random slopes reflect family-specific relationships between distance measures and L2 proficiency. We chose to include by-family random slopes as it may a priori be expected that family biases moderate the relation between morphological distance and proficiency. This theoretical motivation of the random effect structure avoids overfitting to a particular sample (Barr, Levy, Scheepers, & Tily, 2013), in this case a selection of languages from multiple families.

In all, we make use of a LMEM with morphological distance as a language level fixed predictor and random intercepts and slopes across families (Gelman & Hill, 2006). The model has six parameters, comprising three variance components (variance across families, languages, and random slopes), one covariance coefficient (between slopes and families), one fixed effect, and an overall intercept. Adding another distance measure to this model (morphological or lexical) involves estimation of an additional random slope and a more complex covariance structure (random intercepts x distance measure 1, random intercepts x distance measure 2, distance measure 1 x distance measure 2).

Results

Feature Patterns

A distinction between increasing and decreasing complexity is unnecessary if it does not lead to a better explanation of L2 learnability differences than plain similarity does. Increasing and decreasing complexity together should explain at least as much variance in proficiency scores as similarity alone. In addition, if learning additional inflectional morphology is relatively hard for L2 learners, increasing complexity should match the similarity effect better than decreasing complexity does.

Table 2 shows the correlations between adjusted proficiency scores and measures of similarity, increasing complexity, and decreasing complexity for each of the 28 features. For morphological similarity, all eight significant correlations are positive, ranging between .31 and .68, meaning that a structurally different value in the L1 is often associated with a lower proficiency score. We assume that the negative non-significant values reflect sample fluctuations. Table 2 includes the number of languages for which information was available as well, a number that varies between 23 and 53. The varying numbers have an impact on the correlations observed, but it is hard to tell what the precise effects are. The global pattern obviously is that differences lead to impediments, although this is not a consistent finding for all features. Features spread in their effect on proficiency.

This overall view is confirmed for the features with increasing complexity. Some significant correlations are even a bit higher than their similarity counterparts are. Seven correlations are significant, a subset of the eight significant similarity correlations (the exception is feature number 57, coding of possessives). Table 2 shows that increasing complexity captures most of the effects found for similarity. The feature patterns for increasing complexity are highlighted in more detail in Table 3. The correlations are ordered from high to low to illustrate which features affect proficiency the most. The patterns on each row are ordered from lower to higher complexity, e.g. no past tense < past tense. As Dutch has a past tense, languages with no past tense have a lower morphological complexity than Dutch. The positive correlations (14 out of 16) indicate that L1s with a relatively low proficiency score are likely to be less complex than Dutch. Seven out of sixteen correlations are significant and positive, meaning that they are in agreement with the observed feature value orderings in Lupyan and Dale (2010). Two feature patterns have a negative correlation, but in both cases, the correlations are non-significant. Negative correlations can arise because of statistical fluctuation due to the current selection of L1s in the present study.

Table 2. 28 WALS feature numbers, the number of languages with the respective feature available, and correlations between adjusted proficiency scores and distance measures. A blank cell indicates that no language is either more or less complex than Dutch for that feature.

WALS No.	Languages	Similarity	Increase	Decrease
100	30	.68***	.68***	
102	30	.59***	.68***	-.05
29	25	.51**	.71***	-.23
66	39	.43**	.43**	
112	53	.36**	.33*	.09
57	37	.35*		.35*
92	46	.34*	.34*	
26	51	.31*	.31*	
73	41	.29		.29
74	45	.26		.26
20	20	.26	.26	
75	44	.23		.23
22	23	.22	.33	.02
67	39	.19		.19
76	44	.18		.18
41	30	.16	-.32	.36*
65	39	.14	.14	
77	32	.12	.12	
36	37	.09		.09
28	25	.06	.06	
70	51	.04	.04	
59	24	.03		.03
38	40	.00	-.05	.14
49	44	-.28		-.28
98	23	-.22		-.22
48	30	-.13		-.13
37	43	-.09	.06	-.17
101	46	-.06		-.06

*Signif. codes: ***: p < .001, **: p < .01, *: p < .05*

Table 3. The feature hierarchy orders features with the highest impact from high to low. Impact is based on Pearson correlations between predicted and observed differences in morphological complexity between the L1s and Dutch (L2). The patterns in the first column point out which feature values are considered less complex (<) than the value of Dutch (always in last position).

Short Description (WALS no.), tested pattern	r
1. Syncretism in Verbal Person/Number Marking (29), none < syncretic	.71
2. Alignment of Verbal Person Marking (100), neutral (absent) < accusative	.68
3. Person Marking on Verbs (102), no person marking < agent only	.68
4. Past Tense (66), no past tense < past tense	.43
5. Polar Question Coding (92), question particle < no question particle	.34
6. Coding of Negation (112), word/affix/double < negative particle	.33
7. Inflectional Synthesis of the Verb (22), 0-1 < 2-3 categories per word	.33
8. Inflectional Morphology (26), little affixation < strongly suffixing	.31
9. Fusion of Inflectional Formatives (20), isolating < concatenating	.26
10. Perfective/Imperfective (65), no grammatical marking < grammatical marking	.14
11. Coding of Evidentiality (77), no evidential < indirect only	.12
12. Case Syncretism (28), no case marking/core and non-core cases < core only	.06

13. Definite Articles (37), no articles/demonstrative word < word distinct from demonstrative	.06
14. Morphological Imperative (70), no imperatives/singular/plural < 2nd person number-neutral	.04
15. Indefinite Articles (38), no articles/indefinite word same as ‘one’ < indefinite word distinct from ‘one’	-.05
16. Distance Distinctions in Demonstratives (41), no distance contrast < two-way contrast	-.32

What may we expect for decreasing morphological complexity? Learning to make use of less complex morphology in an L2 seems less difficult than learning to make use of more complex morphology. If this were not the case, linguistic structures of exoteric languages should be characterized by less complex morphology, which is not true. Acquiring a language with less complex morphology than is present in the learner’s native language should be easier than acquiring a language with more complex morphology. The implication is that decreasing complexity should correlate less strongly, if at all, with proficiency scores than increasing complexity.

The third column in Table 2 gives the correlations between proficiency scores and decreasing complexity. For decreasing morphological complexity, correlations for 19 features are available. Two features have a significant positive correlation. The first one, with the highest correlation, comprises distinctions in demonstratives related to distance (no. 41, r = .36, p < .05). The second feature is affixal possessive marking (no. 57, r = .35, p < .05), which Dutch does not have. Unlearning to code possessives thus has a significant positive effect – that is, it makes it harder to acquire L2 Dutch. Furthermore, although the other patterns are non-significant by themselves, together they may still have an effect: the seven negatively correlating patterns could suggest that a decrease in complexity is beneficial to learning, instead of adding difficulty. In all, only a few significantly correlating feature patterns based on decreasing morphological complexity are found. This strengthens the evidence for the importance of increasing complexity in L2 learning.

Combining Feature Patterns

By combining feature patterns, a distance measure can be developed to assess the general effect of the complete set of relevant features. Several methods exist to optimize the weighting of the features involved. We decided to avoid any suggestion of maximizing our results by taking a mechanical approach based on the sample correlations we found in our data. To this end, we computed overall scores by weighting all relevant features by their correlations. We only included languages for which more than five feature values were available, reducing the subset of languages to 49. The three resulting distance measures for similarity, increasing complexity, and decreasing complexity give the distances from either Indo-European or non-Indo-European L1s to Dutch. For example, the maximum observed increasing complexity score before dividing is 4.289 for Vietnamese, which is the sum of all 16 correlations in Table 2 with four exceptions: one missing value and three equally complex feature values. Vietnamese has articles, makes perfective/imperfective distinctions, and has distance distinctions in demonstratives. In all other feature correlations, Vietnamese is less complex than Dutch. Dividing the weighted sum by the number of available features (15), Vietnamese gets a score of .286 for increasing morphological complexity. The larger the distance score, the more complex Dutch is as compared to an L1.

Having computed the three distance measures, we find that similarity and proficiency are strongly correlated (r = -.77, $p < .0001$), and similarity is more strongly correlated with increasing complexity (r = .78, $p < .0001$) than with decreasing complexity (r = .60, $p < .0001$). Increasing complexity correlates more strongly with proficiency (r = -.67, $p < .0001$) than decreasing complexity (r = -.45, $p < .005$). Decreasing and increasing complexity are not significantly correlated (r = .23, p = .11). Figure 5 shows a scatterplot of adjusted L2 proficiency scores and increasing complexity from the L1 perspective.

"0" on the x-axis means that an L1 has more or exactly the same degree of morphological complexity as compared to Dutch. This is why most Germanic languages are situated here. Further to the right on the x-axis are L1s with a lower degree of morphological complexity. These languages are mostly not related to Dutch, like Vietnamese. Many Indo-European languages are clustered together. The grey line is a linear fit to the adjusted proficiency scores with 95% confidence intervals added. The linear fit explains a substantial amount (45%) of the variance in proficiency.

We compare maximum LMEMs to exactly the same model without the respective predictor added (a null model) by means of likelihood ratio tests. A maximal LMEM is a mixed effects model with random intercepts and slopes (Barr et al., 2013). There is a significant effect of similarity on proficiency ($\chi^2(1) = 6.86$, $p < .01$). When replacing similarity by increasing complexity, the effect is still significant ($\chi^2(1) = 5.02$, $p < .05$). However, when replacing similarity with decreasing complexity, the effect is not significant anymore ($\chi^2(1) = 0.28$, $p = .60$).

A maximal LMEM with similarity as a fixed effect is not significantly more likely than a maximal LMEM with increasing complexity as fixed effect (evidence ratio of 6.9). An evidence ratio (Spiess, 2013) of 10 or more indicates strong evidence. This means that increasing complexity explains the same amount of variance in proficiency scores as similarity. On the other hand, decreasing complexity provides the least evidence: the similarity model is 34,454.5 times more likely than a decreasing complexity model. Combining increasing and decreasing complexity does not lead to a better model than similarity alone (evidence ratio of 1.9). We conclude that cross-linguistic morphological similarity effects seem to be largely built up from the degree of increasing morphological complexity in the L2.

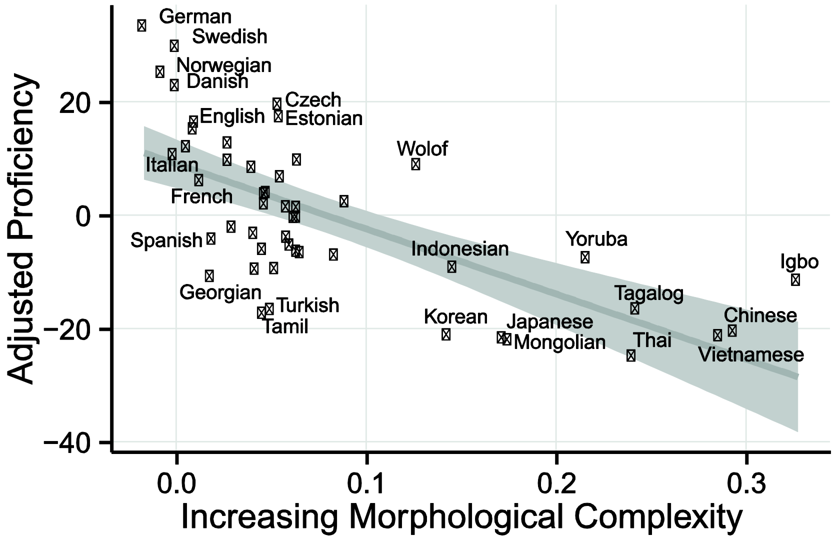


Figure 5. The relationship between speaking proficiency scores and a weighted sum of features as based on the features that are less complex in the L1 than in Dutch. On the y-axis, a score of zero is the average test score across L1s; on the x-axis, zero indicates a strictly equal or higher morphological similarity.

Contrasting increasing with decreasing complexity, we find that a LMEM model with increasing complexity is significantly more likely than a decreasing complexity model (evidence ratio of 4,977.3). Adding increasing complexity to a model containing decreasing complexity (and random slopes for increasing complexity) improves the model significantly ($\chi^2(1)= 5.36$, $p < .05$), whereas adding decreasing complexity to a model containing increasing complexity does not ($\chi^2(1)= .82$, $p = .366$). The distance measure based on increasing morphological complexity seems to overshadow and even nullify decreasing complexity.

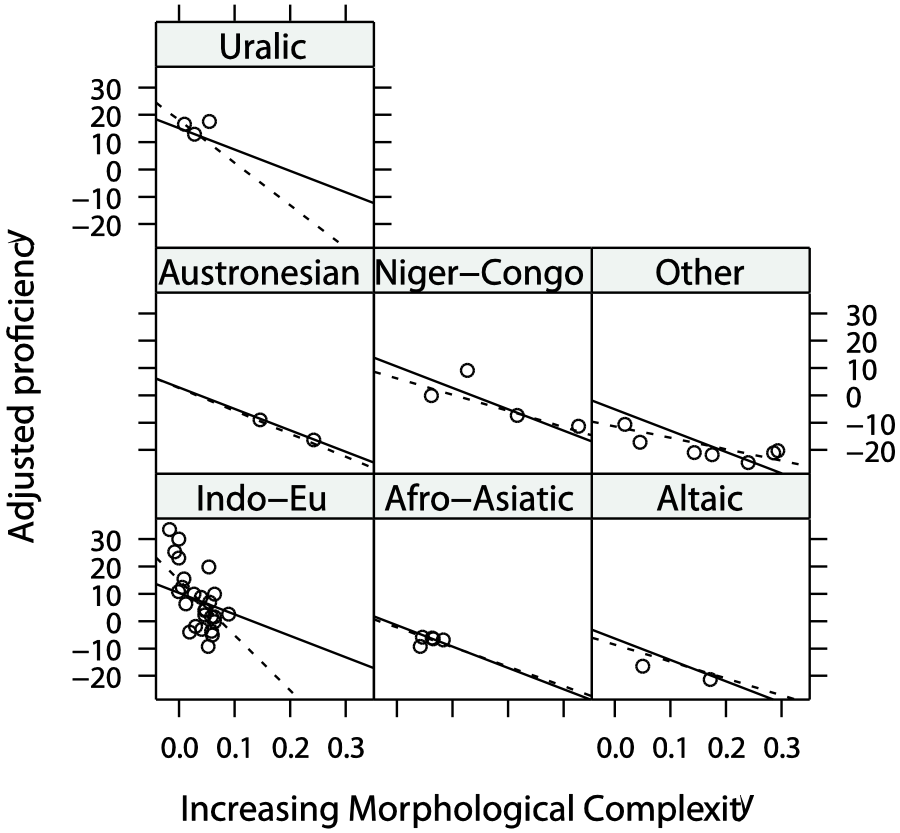


Figure 6. An account of between-family variation in the relation between L2 speaking proficiency scores and features that are less complex in the L1. The panel with the category "Other" contains L1s from families with one or two languages available in our sample. The solid lines are based on a random-intercept only model and the dotted lines on a random-slope random-intercept model.

Family Bias

The relation between complexity and proficiency may be moderated by a family bias. The LMEMs reported on thus far include estimated family-specific random intercepts and slopes for each language family with two or more languages available. The family-specific slopes are shown as dotted lines in Figure 6. The slope of the solid line, which is constant across families, is taken from a model without random slopes. The random slopes are consistently positive across families. The effect is relatively strong for Indo-European and Uralic L1s, as the dotted line for the Indo-European and Uralic families are steeper than the solid slope. This indicates a bias for a relatively high degree of similar morphology, which is unsurprising as Dutch is Indo-European itself. The steep slope for Uralic may have been affected by the Indo-European estimate, as only three Uralic languages are available. For the other three families there is a lower by-family adjustment to the slope, indicating that, within these families, a larger complexity difference is needed for the same difference in adjusted proficiency. Indo-European and Uralic L1s also have a higher random intercept than other L1s, meaning that they are expected to perform better irrespective of their morphological complexity. However, the effect of morphological complexity is robust across random differences between language families.

The relation between complexity and proficiency varies across language families, as the variance across neither the random slopes nor intercepts is zero. Figure 6 shows that the intercepts differ along the y-axis and the slopes differ in steepness. However, the estimated random slopes are highly (but not perfectly) collinear with the estimated random intercepts (r = -.805). The random slope for Uralic languages is less steep than predicted by the random intercepts alone, and the slope for Niger-Congo languages is steeper than predicted by the random intercepts alone. The 95% confidence intervals for the random intercepts of Uralic, Indo-European, and the “Other” category (families with only one or two languages available, see Figure 6) do not contain 0, indicating that these random intercepts are different from the fixed intercept. Similarly, the 95% confidence intervals for the random slopes of Indo-European and the “Other” category do not contain 0, indicating that these slopes are different from the fixed slope. The family bias is not strong enough (in the present set of languages) to reverse directions of any of the random slopes. In a different or larger sample of languages, family bias may play a more critical role.

Lexical Distance

The previous paragraphs show that morphological similarity and increasing complexity are closely related. However, it is still unclear whether the observed effect of increasing complexity cannot be reduced to a lexical distance effect. In other words, we need to investigate whether increasing complexity explains some variance that is not explained by the lexical distance model alone.

Lexical distances are a successful model of similarity effects in L2 learnability (see Section 2). Here, we use lexical distances as computed by summing over inferred branch lengths from an Indo-European language family tree (Gray & Atkinson, 2003). Missing values are replaced with inferred distances from the ASJP tree (Wichmann, Holman, et al., 2010), as calculated by applying the Levenshtein-based LDND distance measure to version 13 of the ASJP database (Schepens et al., 2013b). The set of 49 L1s contains 26 Indo-European L1s. Correlating lexical distances with proficiency scores reveals that the lexical distances are more strongly correlated with proficiency (r = -.80, p < .0001) than similarity (r = -.65, p < .001), increasing complexity (r = -.68, p < .0001), and decreasing complexity (r = -.15, p = .4588).

Within Indo-European languages, lexical distance (F(1) = 43.71, p < .0001), similarity (F(1) = 17.32, p < .0001) and increasing complexity (F(1) = 20.37, p < .0001) are significant predictors of L2 proficiency scores, while decreasing complexity is not (F(1) = 0.57, p = .4588). Lexical distance is a better model than all of the three morphological measures (evidence ratios for similarity: 613.4, for increasing complexity: 244.1, and for decreasing complexity: 530,273.7). Within Indo-European languages, similarity and increasing complexity models are both equally likely (evidence ratio of 2.5), while both being better than decreasing complexity (evidence ratio for similarity: 864.4, increasing complexity: 2172.5). Lexical distance is a significant addition to all three morphological measures used within this language family (similarity: F(1) = 17.9, p < .0001, increasing complexity: $F(1) = 15.1$, $p < .0001$, decreasing complexity: $F(1) = 40.6$, $p < .0001$), whereas none of the three morphological measures adds significantly to lexical distance (similarity: $F(1) = 1.9$, $p < .1779$, increasing complexity: $F(1) = 2.0$, $p < 0.1726$, decreasing complexity: $F(1) = 0.1$, $p < 0.7662$). Within Indo-European languages, lexical distance is thus a better model than either similarity or increasing complexity. Decreasing complexity is worst.

Lexical distances between non-Indo-European languages and Dutch are not available in Gray and Atkinson (2003). In order to incorporate non-Indo-European languages, we assume that their distance to Dutch is maximal. The maximal lexical distance in our subset of Indo-European languages is the distance of Albanian to Dutch. The correlation of lexical distance with proficiency scores ($r = -.73$, $p < .0001$) is similar to the correlations of similarity and increasing complexity with proficiency scores (similarity: $r = -.77$, $p < .0001$, increasing complexity: $r = -.67$, $p < .0001$, decreasing complexity: $r = -.45$, $p < .0001$). Adding similarity ($\chi^2(1) = 6.05$, $p < .05$) and increasing complexity ($\chi^2(1) = 9.05$, $p < .01$) to a lexical distance model improves model fit significantly, but adding decreasing complexity makes no difference ($\chi^2(1) = 0.68$, $p = .41$). Vice versa, adding lexical distance to either a similarity model ($\chi^2(1) = 1.07$, $p < .302$) or an increasing complexity model ($\chi^2(1) = 2.09$, $p < .148$) does not improve the model significantly, but adding lexical distance to a decreasing model almost reached .05 ($\chi^2(1) = 3.84$, $p = .0501$), indicating that similarity and increasing complexity already account for lexical distance. The best model in a sample including non-Indo-European languages is the lexical distance model, as it is more likely than all three morphological models (evidence ratio for similarity: 218.7, increasing complexity: 1,513.8, decreasing complexity: 7,534,621). We already saw above that there is no strong evidence for favoring the similarity model above increasing complexity (evidence ratio of 6.9).

We conclude that adding non-Indo-European languages enhances the role of increasing morphological complexity in explaining L2 learnability differences. Lexical differences can no longer account for distances between Indo-European (Dutch) and non-Indo-European languages. This outcome strengthens the pivotal role of morphology. The effect of increasing morphological complexity is also seen within the Indo-European language family, as indicated by the high correlation between lexical and morphological distance for the Indo-European languages.

Discussion and Conclusion

This study investigated the relation between proficiency measures of adult language learning and cross-linguistic differences in morphological similarity and complexity between 49 different L1s and L2 Dutch. Most of the morphological complexity patterns observed by Lupyan and Dale (2010) were also present across L2 learners of Dutch. To our knowledge, no other study investigates systematically across a large number of languages to what extent morphological similarity and complexity determine L2 proficiency. We used the notion of L2 learnability to capture L1 properties that co-determine adult L2 learning. Our measures were three morphological measures based on similarity, increasing complexity and decreasing complexity. The study employed L2 speaking proficiency scores as a new type of data in the study of cultural-evolutionary mechanisms in language structure (Nettle, 2012). In the present section, the results are discussed with respect to the relation between morphological complexity and L2 learnability, with respect to variation across lineages, and with respect to other measures of linguistic differences.

First, morphological similarity correlated significantly with proficiency in 8 out of 28 features. Seven out of these eight correlations are a result of increasing complexity and one correlation is a result of decreasing complexity. For these seven features, L2 learnability is lower, the less morphologically complex the L1 is compared to the L2. See Tables 2 and 3 above for the feature-specific correlations.

An overall measure of morphological similarity as computed by combining feature-specific correlations yields a correlation of -.77 ($p <$

.001). An overall measure of increasing morphological complexity correlates more strongly with L2 Dutch proficiency scores ($r = -.67$, $p < .001$) than decreasing complexity does ($r = -.45$, $p < .005$). We did not try to optimize these results, for instance by excluding simpler or more complex features of Dutch that correlate with high L2 Dutch proficiency. We wanted to include the whole set to test if morphology has an overall effect on L2 learnability. The individual outcomes for the separate features reflect the overall pattern with similarity and increasing complexity as stronger effects, decreasing complexity being the weaker component. The outcomes provide confirmatory evidence for the validity of the cross-linguistic patterns of morphological complexity that Lupyan and Dale (2010) observed.

The overall outcomes turn out to be robust when controlling for language family biases with LMEMs: an increasing complexity model is 4,977.3 times more likely than a decreasing complexity model. Replacing similarity with increasing complexity (neglecting all differences due to decreasing complexity) does not result in worse model fit. Thus, learning complex morphology seems to be more difficult for adult language learners with less morphologically complex L1s. The finding is consistent with conclusions from psycholinguistic studies (Ionin & Wexler, 2002; Lardiere, 1998; Mitchell & Myles, 2004), e.g., in terms of use of articles, case systems, past/future tense, etc. The effect of morphological complexity in particular remained consistently present and constant when controlling for family-specific biases.

What is the relationship between the morphological distance measures and lexical distance? A comparison between increasing complexity and lexical distance suggests that increasing complexity adds independent explanatory value to a lexical account of L2 learnability. Adding increasing complexity to a lexical distance model improved model fit significantly ($\chi^2(1) = 9.05$, $p < .01$). The complexity measure is more versatile than the lexical distance measure because the morphological measure can be employed for all language families. Again, substituting complexity for similarity does not result in a better model fit. It was concluded that morphological and lexical distance measures are closely related.

Does the way the L1 impacts learning an L2 explain cultural-evolutionary mechanisms of language variation and change? In a social setting, dialect structures are influenced by incomplete transmission due to adult learning (Labov, 1972). Quantitative diachronic investigations of the role of L2 learning, however, suffer from a lack of data to determine the mechanisms in more detail: “The results need more sophisticated multivariate and comparative analysis, and perhaps most pressingly, the cultural-evolutionary mechanisms involved need to be isolated and identified” (Nettle, 2012, p. 1835). Our findings can be helpful. They indicate that transmission of complex morphology in adults is hampered, which is in line with experimental and longitudinal studies of adult L2 learning of morphosyntax (Birdsong & Molis, 2001; Flege, Yeni-Komshian, et al., 1999). Moreover, adult L2 learning depends on the complexity of a learner’s L1 and the complexity of the target language.

Historical linguists have argued that some languages are morphologically more complex than others because of cultural-evolutionary mechanisms that accommodate adult language learning (McWhorter, 2011; Trudgill, 2011). It may be the case that languages gradually adapt to the common cultural practice of speaking foreign languages, similarly to linguistic adaptation to growth of literacy (Levinson & Gray, 2012). Interestingly, the impact of increasing morphological complexity is not equal across all morphological features. Our outcomes provide a rank order or hierarchy of feature impact. Morphological differences low in the hierarchy of impact may be more accessible for transfer, and features higher up in the hierarchy may be more likely to cause substrate effects. The L2 literature offers evidence that supports this hierarchy. For example, Oldenkamp (2013) concludes that verbal inflection of Arabic, Chinese and Turkish learners of Dutch influences production of L2 Dutch constructions with verbal inflection.

The concept of L2 learnability overlaps with cross-linguistic influence (CLI), but both are useful for different theoretical discussions. L2 learnability is a scale measuring the extent to which L2 proficiency depends on the L1 across learners. At the level of the learner, the L2 literature focuses on differences between L2 and L1 proficiency, e.g. differences due to critical period, CLI, and transfer effects. At the level of language structure, we are interested in whether one language is easier to learn as an L2 than another is. Some L1s are more complex than others are (e.g. with respect to morphology), which may affect how difficult it really is to learn an L2 with a complex morphology. We find that this is, on average, the case across the different L1s included in our study. Therefore, in a general sense, morphological complexity plays a decisive role, although the extent to which it determines L2 learnability depends on the relative differences between L1 and L2 morphological complexity. In addition, the concept of L2 learnability entails that L1 acquisition has been completed before the onset of L2 learning. In other words, L2 learnability applies to successive language learning.

The data that are used here provide a unique possibility to assess, on a large scale, quantitative effects of structural differences between L1s. The Dutch-specific proficiency scores are affected by demographic and geographical factors. In a pre-analysis, we controlled for several individual differences. Furthermore, we expect that our findings for L2 Dutch will also play a role in studying the acquisition of other second languages. L2 learnability is not a symmetric notion for all pairs of languages: it depends on which language is the L1 and which language is the L2. It is an empirical question to what extent the internal feature weights may need to be reconfigured for testing on different L1s or L2s. Including additional L2s will give us data to investigate in more detail which morphological features stand out as complex and which do not, and give us a better insight into the complexity of linguistic structures.

A large sample of many linguistic features from a wide variety of languages is provided by WALS. Lupyan and Dale (2010) used a broad selection of features to assess the correlation between population size and the use of morphological vs. lexical encoding strategies. The present study assesses whether Dutch L2 learnability is associated with these same observed feature patterns, as ranging from lexical to morphological. As it turns out, the feature patterns that Lupyan and Dale observed correlate strongly with L2 learnability with only a few exceptions. Although most of the observed patterns reported in Lupyan and Dale indicate high complexity in small languages, plurality was one of the features in their study that did not follow this trend. However, in our study, we found that morphological plurality marking actually is harder to learn in an L2. These conflicting findings may be due to differences in the subsets of languages between the two studies. With respect to distance distinctions (no. 41) and coding of possessives (no. 57), the two features for which our study showed decreasing, rather than increasing complexity to be harder to learn, the subset of languages might have played a role as well. It would be good to investigate the interesting differences from the Lupyan and Dale observations in future work.

LMEMs are a conservative way of modeling lineage and language-specific factors that affect whether or not a language is more or less complex than Dutch morphologically. On average, even after adjusting for random between-family variation, the features with increasing morphological complexity still correlate strongly with L2 learnability. It remains an open question to what extent the variation in specific lineages supports this claim. The data available for the learnability of Dutch as an L2 and the structural configurations of the L1s of the learners together provide a large-scale quantitative source of evidence for the hypothesis that morphological complexity across languages may be constrained by adult language learning over longer periods.

Chapter 4

Learning New Sounds: A Phonological Account of L2 Learnability

Acknowledgments

This chapter comes forth from a research visit to the HLP / Jaeger lab at the University of Rochester. The chapter greatly benefitted from discussion in Rochester as well as in Nijmegen. Besides J.J. Schepens, T.F. Jaeger and R. Van Hout also contributed as co-authors to the manuscript.

Abstract

Understanding the linguistic factors that facilitate or impede second language learning can shed light on the underlying mechanisms of transfer in additional language learning. Based on a large database of 50,000 adult second language (L2) learners of Dutch from 62 native language (L1) backgrounds, we investigated how differences between the phonologies of the L1s of the learners and Dutch affect L2 speaking proficiency. Specifically, we investigate if L1 impact is limited to speech sounds or also includes distinctive features. We find that higher similarity between sound inventories facilitates L2 learnability. This effect is stronger, the more similar the new sounds are to their most similar L1 sound neighbors. These results provide support for accounts of L2 learning that stress the importance of transfer from L1 to L2. Critically, this transfer goes beyond simple transfer of sound categories at the earliest stages of L2 acquisition. Instead, our data suggest that L2 learners a) continue to engage in cross-linguistic inference throughout the entire period of L2 acquisition and use and b) that they do so at multiple levels of linguistic representation, including subphonemic details.

Keywords: Transfer, Phonological Similarity, Phonological Distance, L2 Learnability, Sound Inventory, Distinctive Feature

Learning New Sounds: A Phonological Account of L2 Learnability

Introduction

Adult language learners face many challenges in acquiring a second language (L2). The challenges include the acquisition of new syntactic structures, lexical items, morphological paradigms, and phonological properties. Given the magnitude of this enterprise, it is hardly surprising that many adult language learners fail to converge on native-like proficiency, even after years of exposure to the new language. Understanding the factors that affect L2 learning and the level of proficiency in a facilitative or prohibitive way is imperative to cognitive scientists and to linguists because language learning holds the key to a better understanding of the mechanisms underlying language use and the mechanisms involved in learning more generally.

Pronunciation problems and a non-native accent are amongst the most prominent characteristics of L2 speech that native speakers observe (Strange & Shafer, 2008). L2 accents induce inferences about, for example, the L2 learner's general intelligence, social status, and attractiveness (Munro, 2008), also affecting their economic mobility (Bleakley & Chin, 2010; Johnson, 2014; Lopez, 1999; Saiz & Zoido, 2002). L1 sounds and sound patterns keep remaining active in the new additional language, for instance in the recognition of new words (Cutler, 2012). This chapter investigates the impact of the distance between L1 and L2 sound inventories and their distinctive features on L2 learning difficulty: Our central claim is that the larger the distance, the larger the degree of difficulty to learn the L2 and, as a consequence, the lower the degree of L2 learnability.

Research has revealed a number of factors contributing to L2 proficiency including age of acquisition (Lenneberg, 1967), duration of exposure (Pica, 1983), and individual differences in language learning aptitude (Schumann et al., 2004).[5] The discussion currently centers around the question what transfers from a learner's language background (Foley & Flynn, 2013; Kellerman & Sharwood Smith, 1986), the hypothesized non-linearity of age effects (Birdsong, 2014), the importance of education and literacy (Hakuta et al., 2003; Huettig, in press), and general advantages of bilingualism (Bialystok, 2013; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009).

It has long been assumed that learners make use of similarities between the target L2 and languages previously learned to transfer L1 knowledge when learning an L2 (Lado, 1957; Weinreich, 1963). Indeed, a large number of studies has confirmed this idea, drawing on a variety of methodological approaches (Cenoz et al., 2001; Chiswick & Miller, 2005; De Angelis & Dewaele, 2011; De Angelis & Selinker, 2001; Foley & Flynn, 2013; Gass & Selinker, 1992; Ionin & Montrul, 2010; Isphording & Otten, 2013; Jarvis & Pavlenko, 2008; Kellerman & Sharwood Smith, 1986; Major, 2008; Odlin, 1989; Ringbom, 2007). These studies have shown that similarities between previously learned languages and additional languages facilitate language learning in the domains of lexicon, syntax, and phonology.

Studies of language transfer have found that even small L1-L2 sound differences may inhibit successful L2 learning (Best, 1995; C. Brown, 2000; Flege, 1993; Kuhl, 1991). For example, when adult Italians learn the English sound /eɪ/ (as in *play* or *lane*), they seem to assimilate /eɪ/ to the phonologically similar L1 category /e/ (as in *bed*) (Piske, Flege, MacKay, & Meador, 2002). While previous research has provided ample evidence for the interference caused by L1-L2 differences (Flege, 2003; Flege, MacKay, & Meador, 1999; Piske et al., 2002), the precise nature of L1-L2 interactions remains unclear.

[5] Although all of these factors have been found to correlate with L2 proficiency, many if not all of these factors tend to be highly correlated, making it hard to tease apart their respective contributions. Additionally, although research has identified many intrinsic and extrinsic factors (see e.g. Moyer, 2013, p. 83), we know of no studies that assess the relative contributions of all of these factors simultaneously. Consequently, it is not surprising that the role of many of the above factors in language learning remains a matter of debate.

Consequently, Flege (2003) concluded: “It will be necessary to study a wide range of L1-L2 pairs and L2 speech sounds in order to draw general conclusions regarding the nature of constraints, if any, on L2 speech learning.”

Present Study

The present study aims to contribute to this goal. Drawing on a large database of L2 Dutch learners from many different language backgrounds, we test whether phonological similarity between the L1 and L2 sound inventories facilitates L2 learnability. To investigate L2 learning difficulty of L1-L2 phonological differences, we aim to give a phonological account of L2 learnability, which we define as the learning difficulty for a given L1. For example, not all L1s use vowel length to distinguish short and long vowels, leading to different L2 learning challenges across L1s. We test the hypothesis that the L2 learnability of a language’s phonology depends on the phonological structure of previously learned languages. Here, we restrict our investigation to the sound categories of a language (i.e., the sound inventories) and the phonological features contrasting these categories (distinctive features). Other aspects of L2 phonological learning include the phonotactic structures (e.g., syllables) and suprasegmental characteristics, which we leave to future work.

Given our focus on sound inventories, our **main hypothesis** is that the phonological similarity between the sounds in the L1 and the L2 sound inventories accounts for variation in L2 learning across L1s. If we find that distance measures between the L1 and L2 sound inventories affects L2 learning, this would mean that theories of L2 learning need to be able to account for the way learners can benefit from similarities (small distances) or are impeded by dissimilarities (large distances).

In addition, we investigate whether the L1 distinctive features are relevant for L2 learning. Do L1 sounds impact how additional L2 sounds are learned (Flege, 1993) or is it primarily the distinctive features of the L1 sounds that determine transfer (C. Brown, 1998, 2000)? This way, our **second hypothesis** is that transfer effects are not limited to similarities and differences in the acoustic and articulatory make-up of sounds, but that learners are subject in their learning behavior to similarities and differences in the featural representations of sounds as well. The features may perhaps help to understand what makes new sounds so hard to learn. Our second hypothesis resembles our main one, but is now applied to features, distances between L1 and L2 being defined on the basis of the number of features involved.

We assess the influence of L1 phonology on L2 learning across 62 different L1s. Our data comes from the “Dutch as a Second Language” state exam (henceforth STEX). STEX is highly heterogeneous, with learners from a wide range of ages, educational and language backgrounds, and exposure times to Dutch. To control for other factors known to affect L2 learning, we employ mixed effects regression modeling. The approach taken here thus complements experiments that focus on two to a handful of languages, while aiming to hold constant other variables.

Recent typological overviews (Dryer & Haspelmath, 2011; Maddieson, 2011a, 2011b; Moran et al., 2014) document the degree of phonological diversity across languages, which opens the way for measuring phonological differences across many different languages. We chose to adopt a quantitative approach by comparing the sound inventories of 62 languages (the L1s) to Dutch (the L2). Dutch has 38 sounds excluding the schwa (Luyckx, Kloots, Coussé, & Gillis, 2007). The Dutch vowel inventory contains 5 lax vowels (ɑ, ɔ, ɛ, ɪ, ʏ), 7 tense vowels (aː, eː, i, oː, øː, u, y), and 3 diphthongs (ɔu, ɛi, œy). The Dutch consonant inventory contains 6 plosives (b, d, k, p, t, ʔ), 9 fricatives (f, ɣ, ɦ, s, ʃ, v, χ, z, ʒ), 2 glides (j, w), 2 liquids (l, r), and 4 nasals (m, n, ɲ, ŋ).

We first present our methods in more detail. This includes the STEX data, our approach to quantifying L1-L2 phonology, and a summary of the statistical procedure. Subsequently, we present Study 1, which assesses the effect of L1-L2 similarity in sound inventories.

Study 2 assesses the effect of L1-L2 similarity in distinctive features. Study 3 compares the resulting distance measures to each other, as well as to lexical and morphological distance measures. The latter distances are essential in providing evidence that phonological distances have their own, autonomous contribution in explaining L2 learnability. We end with a general discussion and conclusions.

Methods

Data

STEX contains L2 proficiency scores for 50,235 learners of Dutch from 74 L1 backgrounds. STEX is tailored to higher education and passing it is a requirement for admittance to a Dutch university. L2 speaking, writing, listening, and reading proficiency scores are available for most participants. Here we focus on the speaking scores, as we suspect L2 phonological learning to be clearly –though certainly not exclusively—reflected in speaking proficiency. The STEX speaking exam consists of 14 tasks that have to be completed in 30 minutes. Participants are requested to provide information, give instructions, and so on. Two independent examiners evaluate the spoken language on both content and correctness according to a formal protocol. The pass level is upper-intermediate, roughly equivalent to the B2 level of the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). More background characteristics of the sample are discussed in Schepens et al. (submitted).

We excluded all languages with fewer than 20 learners in STEX, which leaves us with learners speaking 74 different L1s. We were able to obtain information about the phonological systems of 62 of these L1s (see next section), leaving 48,219 learners for analysis. Following the WALS classification method (Dryer & Haspelmath, 2011), the sample contains languages from 35 different genera and 12 different language families. 33 of the languages are Indo-European, 29 are non-Indo-European. There are eight Niger-Congo, six Afro-Asiatic, four Austronesian, three Altaic, two Uralic, one Dravidian (Tamil), one

Austro-Asiatic (Vietnamese), one Tai-Kadai (Thai) language, and the isolates Japanese and Korean.

For all learners, STEX also provides information about several relevant control variables. These consist of age, gender, educational level, country of birth, length of residence in the Netherlands, age of arrival in the Netherlands, and additional languages learned prior to learning Dutch. We added a measure of the quality of education for the learners' country of birth (see Chapter 2).

Sound Inventories

A sound inventory contains the distinctive segments or speech sounds that build the words of a language. The sound inventory is part of the grammar of a language, e.g. the English sound inventory is described in the grammar of English (Huddleston & Pullum, 2002). These grammars form part of the empirical data that typologists use to study universal trends in the sound inventories of languages across the world (Ladefoged & Maddieson, 1998; Maddieson, 1984). The definition of sound inventories requires the expertise of professional linguists, and these inventories may omit some (e.g. infrequent) variants in a language. For the present chapter, we make use of sound inventories as defined by experts in phonology (Crothers, Lorentz, Sherman, & Vihman, 1979; Lev, Stark, & Chang, 2012; Maddieson, 1984; Moran & Wright, 2009), available in PHOIBLE (Moran et al., 2014). For example, the English phoneme inventory is taken from the Stanford Phonology Archive, which uses information from various sources (Gimson, 1962; Halle, 1973)

To illustrate how it works we look at the differences between the Dutch sound inventory, the sound inventory of a closely related language (English) and the sound inventory of a language that is not related to Dutch (Korean) are shown in Tables 1 and 2.

Table 1. Vowel inventory of Dutch (bold), without diphthongs ɔu, ɛi, œy.

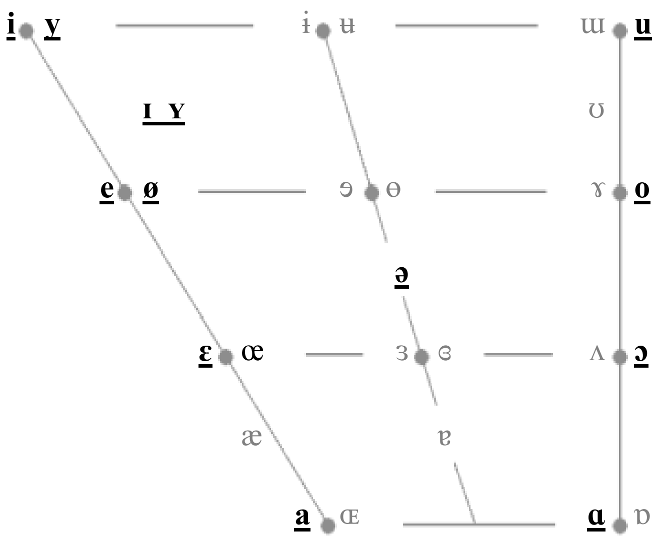


Table 2. The consonant inventory of Dutch (bold), without /w/.

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	**p b**			**t d**		ʈ ɖ	c ɟ	**k** g	q ɢ		**ʔ**
Nasal	**m**	ɱ		**n**		ɳ	**ɲ**	**ŋ**	ɴ		
Trill	ʙ			**r**					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	**f v**	θ ð	**s z**	**ʃ ʒ**	ʂ ʐ	ç ʝ	**x ɣ**	χ ʁ	ħ ʕ	h **ɦ**
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	**j**	ɰ			
Lateral approximant				**l**		ɭ	**ʎ**	ʟ			

PHOIBLE lacked data for a number of L1s. As a result, twelve of the L1s in our sample of STEX were not available in PHOIBLE. Excluding these L1s, left 62 L1s from 48,219 learners for analysis.

Distinctive Features

Distinctive features describe how the sound inventory is built up and how sound differences are defined. Distinctive features are properties of sounds that can be used to characterize speech sounds in terms of the phonological system (Chomsky & Halle, 1968; Jakobson, 1941, 1968). Distinctive features indicate whether a specific movement of oral articulators or the larynx is present (1) or absent (0). For example, the presence of the feature "sonorant" indicates that production of that sound requires continuous airflow in the vocal tract. The presence of "consonantal" requires the (partial) closure of the vocal tract. The presence of "continuant" requires incomplete closure of the vocal tract. The presence of "syllabic" requires the production of a syllable nucleus, and so on and so forth.

Irrespective of whether features are innate or learned early in life, differences in the features required to produce L1 and L2 sounds may lead to persistent difficulties in the acquisition of new sounds later in life. To investigate the role of features, we compare sound inventories based on the distinctive features of their sounds and relate those comparisons to L2 learnability. The number of distinctive features necessary to encode the sound inventory however correlates with the size of the sound inventories themselves, following a certain degree of economic organization (Clements, 2009; Moran & Blasi, 2014). Phonological features also play a role in phonotactic rules (Frisch, Pierrehumbert, & Broe, 2004) and may even play a role in discriminating syntactic categories such as nouns and verbs (Farmer, Christiansen, & Monaghan, 2006). We use a compositional system of distinctive features (Hayes, 2011) for the comparisons of feature representations. As many sounds of the world languages involve alternations for which the featural representation is not fixed as based on Hayes (2011), the feature values of base sounds (e.g. [p]) were supplemented with feature values of components of sounds (e.g.

/aspiration/) to create the encoding of non-base sounds (Moran & Wright, 2009).[6]

Analysis

We use mixed effect regression (D. Bates et al., 2014) to distinguish between individual learner, language, and country level variation in the observed L2 proficiency scores, and to check the consequences of assuming that these levels vary independently at the level of individual proficiency scores. Our null model includes Gender, Age at Arrival, Exposure, Full-time Education, Educational Quality (and their interaction), and random effects for Country of Birth, the L1, the L2, and L1 by L2 (Schepens et al., submitted). Furthermore, we extract the by-L1 adjustments (BLUPS) from our null model in order to be able to report correlation statistics between phonological distance and L2 learnability.

Study 1

[6] We use the extended version of this compositional system, which extends the set of 30 features (Hayes, 2011) into 38 features (Moran & Wright, 2009). This system is able to encode the sound inventories of about 71% of the languages of the world (Moran & Wright, 2009). The features can be grouped into root, laryngeal, supralaryngeal, and place features. The root features are [long], [nasal], [stress], [approximant], [consonantal], [tense] (only applicable when [consonantal] is absent), [sonorant], [delayed release] (only applicable when [sonorant] is absent), [syllabic], and [tone]. The laryngeal features are [fortis], [voice], [constricted glottis], [spread glottis]. The supralaryngeal features are [continuant], [lateral], [tap], [trill]. The place features (as part of the supralaryngeal features) are [labial], [round] (only when [labial] is present), [labiodental], [coronal] and [dorsal]. In addition, [anterior], [distributed], and [strident] are only applicable when [coronal] is present and [front], [low], [back], and [high] are only applicable when [dorsal] is present. Moran & Wright (2009) extended these features with [periodic glottal source] (absent for Dutch fricatives and plosives), [retracted tongue root], [advanced tongue root], [epilaryngeal source], [raised larynx ejective], [lowered larynx implosive], [click] (all never present in Dutch).

Phonological Distance Based on Sound Inventories

What phonological distance measure or measures account for L2 learnability? Traditionally, the degree of underspecification in the L1, the sounds the L2 has that L1 does not have, is brought up as the cause of deviant patterns of variation in the production of new sounds, decreasing the L2 learnability of a target language (Eckman, Elreyes, & Iverson, 2003; Lado, 1957; Major, 2008; Weinreich, 1963). Also, in perception, the listener needs to establish some form of pre-lexical phonological abstractions to represent the new sounds of a target language (McQueen, Cutler, & Norris, 2006). In Study 2, we further define how L2 sounds are underspecified in the L1. Here in Study 1, we define measures that simply count the number of shared sounds, new sounds, and disused sounds (sounds present in the L1, but not in the L2).

Assuming that learners use existing L1 sounds for distinguishing and generalizing to new sounds, we test in Study 1 whether the number of new sounds in an L2 sound inventory explains variability in L2 learnability, and whether shared or disused sounds matter as well.

Measuring Distance between Sound Inventories

A simple distance measure between two inventories may incorporate the increase and decrease in the sound inventory, or increase or decrease only. Measures of the difference between the two sound inventories will be based on three categories of sounds that can be distinguished in comparing two inventories:

1) The shared sounds of the L1 and the L2 (the intersection), see panel 1 in Figure 2
2) The new sounds (the complement of L1 sounds in the L2), see panel 2 in Figure 2
3) The disused sounds (the complement of L2 sounds in the L1), see panel 3 in Figure 2

The first category, the shared sounds, is the complement of the second category, the new sounds. As the L2 is constant in our study, the shared sounds distance is inversely related to the new sounds distance. This means that both measures have an identical explanatory value. In the remainder of this chapter, we therefore focus on the number of new sounds. Distance can be defined from the perspective of the L1, or as an overall difference (different sounds) between the two inventories involved. In addition to the three measures defined above, we define the overall number of different sounds as the sum of the new and the disused sounds. We will use the remaining two categories of sounds (new and disused sounds) as a measure of distance plus the overall difference (new plus disused sounds) and test which of these three measures performs best in explaining L2 learnability.

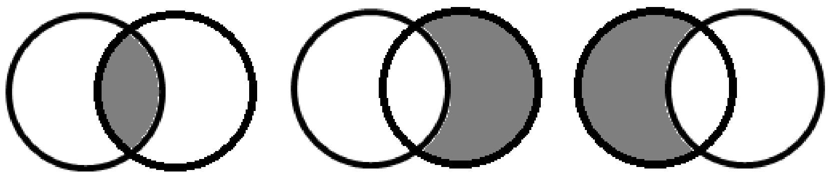


Figure 2. Shared sounds (left), New sounds (middle), Disused sounds (right).

Example

As an example, we give the measures of new sounds, shared sounds, disused sounds, and different sounds with respect to the sound inventories of Korean and English. Both English and Korean have 40 sounds in their inventories, while Dutch has 38 sounds. English learners of Dutch need to learn 19 new sounds, they can make use of 19 shared sounds, they have 21 disused sounds, and there are 40 different sounds in total. Korean learners of Dutch need to learn 22 new sounds (3 more than English learners of Dutch), they can make use of 16 shared sounds (3 less than English learners of Dutch), they have 21 disused sounds (the same as English learners of Dutch), and there 43 different sounds in total.

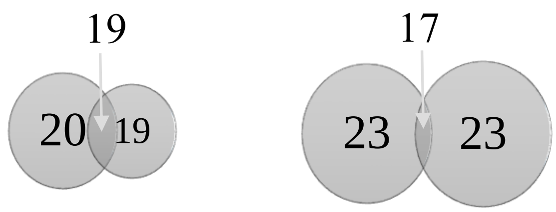


Figure 2. The differences between the sound inventory of Dutch (containing 38 sounds, right circle in both panels) and the sound inventories of English (left circle of the left panel) and Korean learners (left circle of the right panel). The numbers depict the size of the circles *excluding* the overlapping parts. The circles represent the acoustical spaces of the languages. Both English and Korean sound inventories contain 40 sounds (i.e. 21 + 19). Learners of an additional language expand their acoustical space to include both circles.

Results

Above, we defined four inventory-based measures of phonological distance for L2 learners. Here, we test whether these measures explain variation in L2 learnability. The number of new sounds correlates significantly ($r = -.35$, $p < .01$) with speaking proficiency scores, see Figure 3, which we controlled for individual differences, see Methods. A mixed effects regression model (see Methods for a description), shows that for each new L2 sound, a learner needs to deduct 1.007 points from his expected speaking proficiency score. The number of new sounds improved our null model significantly $\chi^2(1) = 6.75$, $p < .05$).

The other three measures (disused sounds, different sounds) did not correlate with speaking proficiency. In contrast to learning new sounds, disused sounds did not correlate with L2 speaking proficiency ($r = -.193$, $p = .13$). The correlation became even weaker after we removed four languages that have many geminates such as Arabic, Amharic, and Hindi ($r = -0.09$, $p = .50$). Geminates are consonants that come in pairs with an audible difference in [length]. Removing these languages did not affect the correlation for the number of new sounds (coefficient and p-value remained identical). The fourth correlation for symmetric difference was non-significant (r = -.09, p = .50). In addition, we tested whether the number of new sounds was dependent on the number of existing sounds by dividing new sounds by the L1 sound inventory size. This also proved to be non-significant (r = -0.22, p = .10).

We conclude that the number of new sounds successfully explains variation in L2 learnability. This suggests that learners benefit from similarity between sound inventories in terms of new sounds.

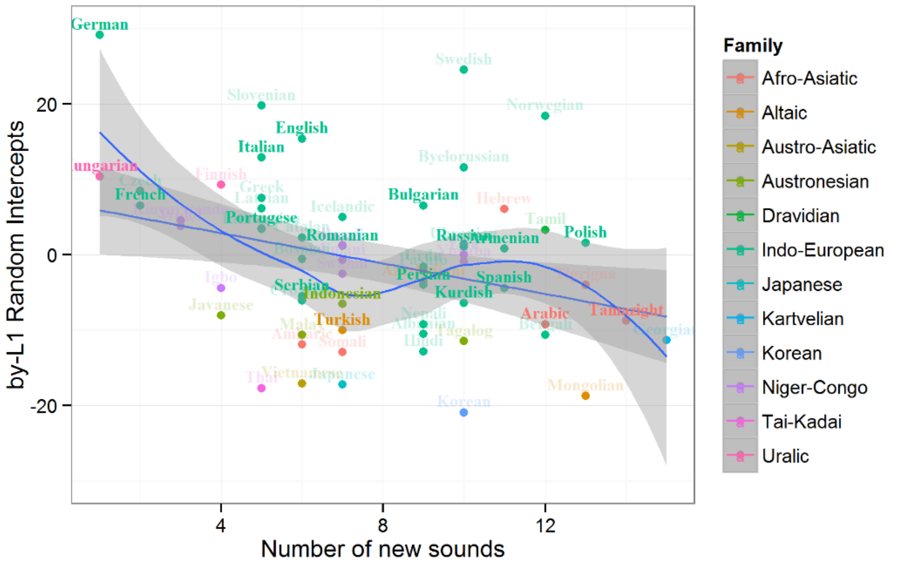


Figure 3. The relation between L2 learnability (measured in terms of by-L1 adjustments controlled for third factors) and the number of new sounds for every L1. The blue lines represent a linear regression and a smoothed fit curve, both with 95% confidence intervals. Indo-European languages generally fall above the regression line, which may indicate lexical and morphological advantages (See Study 3 of the present chapter as well as Chapters 2 and 3).

Study 2

Phonological Distance as based on Distinctive Features

Here, we study the role of distance between new sounds and existing L1 sounds. The effect of distance between sound inventories may potentially be conflated by an underlying effect of distance between sounds themselves. A new sound that uses unfamiliar articulatory features may obstruct an accurate perception or production of the new sound. Learners may consistently keep track of an internal phonetic feature system that they use for the categorization of L2 sounds (C. Brown, 2000). Difference in distinctive features puts a higher learning load on this system by requiring articulators to move in an unfamiliar way. Distant new sounds require the learner to expand his or her internal feature geometry. Thus, phonological transfer may not be limited to sounds, but learners may transfer the featural representations of sounds as well. If that is the case, we should observe that a measure of the distance to new sounds explains L2 learnability beyond the distance between sound inventories. We test whether the similarity with and distance to new sounds explain variation in L2 learnability in Study 2. In order to test hypothesis 2, we need to specify a measure of phonological distance between sounds.

Measures of the similarity between sounds have provided successful explanations of perceptual similarity data. These similarity measures include acoustic (Flege, 1987; Strange, Bohn, Trent, & Nishi, 2004) and articulatory similarity (Bailey & Hahn, 2005; Frisch et al., 2004). Acoustic measures rely on e.g. VOT, formants, while articulatory measures rely on differences in distinctive features of sounds. Such theoretical underpinnings of similarity provide valid models of behavioral measures of similarity such as perceptual similarity judgments (Bradlow & Bent, 2008), confusability judgments (Bailey & Hahn, 2005), or intelligibility data (Gooskens, 2007; Gooskens, Van Heuven, Van Bezooijen, & Pacilly, 2010), as long hypothesized (Shepard, 1987; Tversky, 1977). In addition, typological comparisons (Atkinson, 2011; Moran, McCloy, & Wright, 2012) also consider the importance of the distances between the sounds in sound inventories, i.e. by cross-linguistically comparing the minimal number of distinctive features necessary to describe the sound inventories of the worlds' languages (Moran & Blasi, 2014).

In order to define a distance measure based on distinctive features, we need to specify how L2 sounds relate to L1 sounds. Similarity between two sounds can be measured by counting their shared distinctive features, while distance between two sounds can be measured by counting their different distinctive features (Bailey & Hahn, 2005). When Baily and Hahn correlated these featural measures to perceptual similarity judgments, they found stronger effects (adjusted R^2 = .67 and .62 respectively) than when they correlated discriminability judgments with perceptual similarity (adjusted R^2 = .31). Their finding corroborates the hypothesis that featural distance is characteristic for behavioral measures of similarity, and that different features provide more accurate measures of similarity than shared features. In Study 1, we found an impeding effect of the number of new sounds on L2 learnability, which is a measure of the part of the sound inventory that learners need to add to their L1. At the feature level, the new features of new sounds measure the part of the new sounds of the L2 sound inventory. This way, we expect that new features can explain why learning new sounds impedes L2 learnability.

Measuring Distance between New Sounds and L1 Sounds

Where the intersection and complement of the L1 sound inventory constituted the L2 sound inventory in Study 1, the intersection and the complement of the features of an L1 phonological neighbor to a particular new sound constitute all the features of that new sound in Study 2. We started by counting the number of new features of a new L2 sound. This measure corresponds to the second panel of Figure 2, for each individual sound. We assume that the learners map new sounds onto the L1 sound for which the minimal number of new features are necessary. We sum over the number of new features for each new sound to get the overall new feature distances between the L1 and L2 sound inventories. The resulting distance measure, which has higher values the more different sounds the L2 has, models the hypothesis that learning distant new sounds impedes L2 learnability. Formally, we define the measure of new features as the sum of the minimum complements of the present distinctive features in L1 sounds for all new L2 sounds.

The new features can be determined using the binary feature values of the distinctive features of the two sounds (present / absent). The new features are the present features of the new sound that are not present in the existing sound. It is possible to think of features as absence and presence in such a way that the present features characterize the sounds. Different feature systems may have different geometries of coding the applicability relations between features. It is not exactly clear to what extent the mind encodes the same geometry between features as linguistic feature systems do. A linguistic feature system from a typological perspective may be too elaborate from a cognitive perspective. It is an empirical question whether the distance between present to absent is of equal importance to L2 learning as the distance between present to not applicable, as this depends on the particular feature geometry in the minds of the learners. Because of the issue of distance between presence, absence, and inapplicability, we define another distance measure based on the new features of both present and absent articulatory features of L1 sounds in the new L2 sound, instead of present features only.

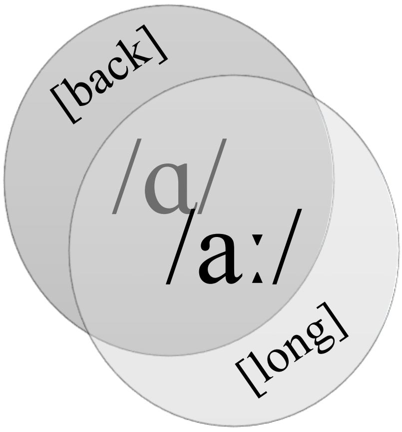


Figure 4. The new feature of /a:/ with respect to /ɑ/ is the articulatory feature [back] and the new feature of /ɑ/ with respect to /a:/ is the articulatory feature [long].

Example

Table 3 shows the different features of the [back] /ɑ/ and the [long] /a:/, as depicted in Figure 4. Both sounds have only one new feature in the other sound. To give another example, Dutch has a [high] [front] [round] vowel /y:/ not present in English and Korean. The Dutch /y:/ has one more feature, i.e. [front], than the Korean and English [high] [back] [round] vowel /u:/ besides the 9 features that it shares. As the Dutch /y:/ is [front], the number of new features compared to /u:/ for both Korean and English is one, see also Table 4. Summing over all Dutch sounds, Dutch has 11 new features compared to English, and 22 compared to Korean, see the x-axis of Figure 5 for the numbers of new features of other L1s.

Table 3. Example of a comparison of the new and shared features for /aː/ relative to /ɑ/.

nr.	Feature	/ɑ/	/ aː/	feature new	feature shared
1	[syllabic]	1	1	0	1
2	[long]	0	1	1	0
3	[sonorant]	1	1	0	1
4	[continuant]	1	1	0	1
5	[approximant]	1	1	0	1
6	[dorsal]	1	1	0	1
7	[low]	1	1	0	1
8	[back]	1	0	0	0
9	[periodic glottal source]	1	1	0	1

Table 4. Examples of sound comparisons.

nld	sound	eng	new features	shared features	sound	kor	new features	shared features
œy	new	uː	1	9	new	y	0	10
p	new	pʰ	0	2	shared	p	0	0
ʀ	new	uː	2	6	new	o	2	6
s	shared	s	0	6	shared	s	0	6
ʃ	shared	ʃ	0	6	new	t̠ʃʰ	1	5
t	new	l	0	3	shared	t	0	3
u	new	uː	0	11	shared	u	0	11
v	shared	v	0	6	new	o	3	3
w	shared	w	0	10	shared	w	0	10
x	shared	x	0	5	new	iː	2	3
ʏ	new	uː	1	9	new	uː	1	9
yː	new	uː	1	9	new	uː	1	9
z	shared	z	0	7	new	sˀ	1	6
ʒ	shared	ʒ	0	7	new	t̠ʃʰ	2	5
ʔ	shared	ʔ	0	1	shared	ʔ	0	1
	disused	iː	0	10	disused	t̠ʃʰ	0	6
	disused	kʰ	0	4	disused	pˀ	0	3

Note: The table shows the number of new features and shared features for a number of sound comparisons to English and Korean phonological neighbors.

Results

Above, we introduced a measure of the number of new features based on the hypothesis that the new feature distances impede L2 learnability. A shift to the level of distinctive features to measure distance is only necessary when it provides a better explanation than a measure of distance at the sound inventory level.

At the feature level, higher L2 learnability significantly correlated with lower numbers of new features ($r = -.47$, $p < .001$, see Figure 5)[7]. The alternative measure that was based on new features with respect to both presence and absence feature values returned a lower (and non-significant) correlation ($r = -.22$, $p < .10$).

The difference between the correlations of new sounds and new features suggests that it may be necessary to explain differences in L2 learnability beyond the level of new sounds at the level of distinctive features. Model comparison shows that the number of new features results in a higher improvement in model fit as compared to the number of new sounds ($\chi^2(1) = 15.00$, $p < .001$ for new features vs. $\chi^2(1) = 6.75$, $p < .05$ for new sounds). The inclusion of random slopes for new features across countries of birth results in a slightly stronger difference in improvement of model fit for both measures.[8] We conclude that the new feature measure results in better fit to the data than the inventory level measure.

[7] In order to assess the robustness of these correlations, we wanted to know whether they remain the same when we exclude learners who are familiar with a language besides their L1 and Dutch. Excluding multilingual learners results in a lower number of observations: The number of languages decreases from 62 to 30 (including only L1s for which we have at least 15 monolingual speakers). The correlation of L2 speaking proficiency with complements remains constant. To overcome the relatively low number of monolingual speakers in our dataset, we additionally selected the group of L3 learners who all report English as their best additional language. This results in a decrease from 62 to 51 languages. The correlation of speaking proficiency with new features again remained stable. In all, the results seem robust to different subsets of the data depending on additional language background.

[8] The model for new sounds did not fully converge.

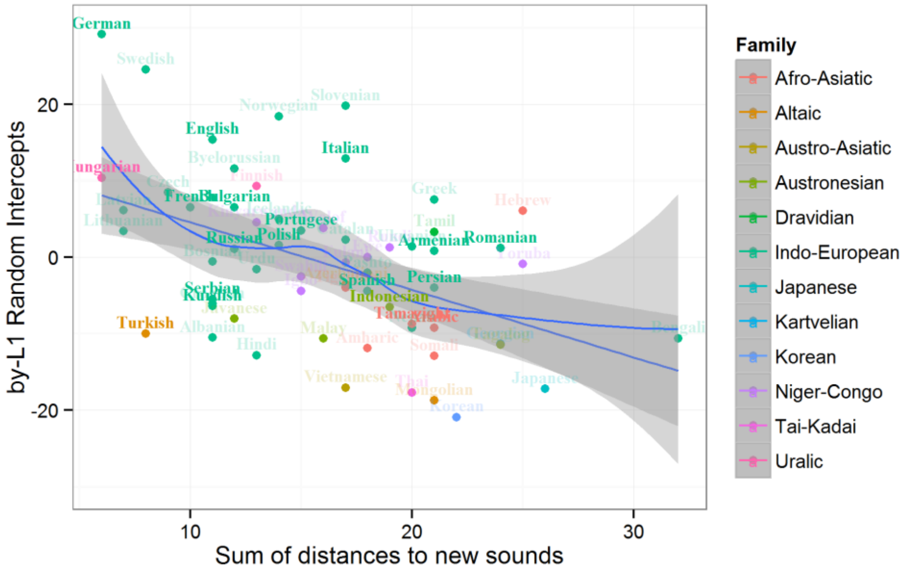


Figure 5. The relation between L2 learnability (measured in terms of by-L1 random intercepts controlled for third factors) and the new features distance (sum of distances to new sounds). The new features distance is measured in terms of the minimal number of new features to new sounds. A high new features distance denotes many new features.[9] As in Figure 3, the blue lines represent a linear regression and a smoothed fit curve, both with 95% confidence intervals. For new feature distances, Indo-European languages fall above the regression line as in Figure 3.

To further assess which model of the three is the best, we compare evidence ratios based on the AIC of each of the three models (Spiess, 2013). The new feature distance measure was significantly

[9] Although Hungarian lacks 13 Dutch sounds, only 6 features of these sounds are missing in the minimally distant sounds of Hungarian. For example, Dutch x (voiceless velar fricative) has [continuant] whereas Hungarian ɟj or ɟj: (voiced palatal affricate) has not. Note that ɟj has [coronal], [distributed], [front], and [periodic glottal source], which do not affect this weight. An L1 Hungarian learner of L2 Dutch (Jani Matyas) confirmed that he felt that Hungarian sounds are similar to Dutch sounds.

more likely than the inventory level measure (1156.907; using evidence ratio's Spiess, 2013). An evidence ratio of 10 indicates strong evidence.

We continue with comparisons of models that include a combination of the measures. The model fits are evaluated for models that include a random effect for country of birth but no random slopes. Adding new features to a model containing new sounds improves the model significantly (χ^2 (1) = 7.95, p < .005), whereas adding new sounds to a model containing new features does not improve the model significantly (χ^2 (1) = 1.19, p = .273). In other words, it is safe to remove new sounds from a model that contains new features, but not vice versa. This indicates that new features explain L2 learnability beyond new sounds.

We conclude that the number of new features successfully explains variation in L2 learnability beyond the number of new sounds. Learners benefit from similarity to new sounds in terms of the new features.

Study 3

Comparing Linguistic Distances

Table 5 shows the correlations between speaking proficiency and the distance measures. The number of new sounds correlates with new features (r = .51), showing the measures give different outcomes.

Lexical and morphological distance already provide good fit to differences in L2 learnability across L1s (Schepens, Van der Slik, & Van Hout, 2013a; Schepens et al., 2013b). We need to be careful when comparing these correlated distance measures. Still, if phonological distance adds to lexical and morphological distances, this would show that the phonological distance contributes to L2 learnability, independently of the other measures.

Table 5. Correlation matrix of L2 learnability (L2), New sounds (1 NS), New features (4), Morphological (Morph.), and Lexical distance.

Measure	L2	New sounds	New features	Morph.
New sounds	.35	x		
New features	.47	.51	x	
Morphological distance	.59	.08	.49	x
Lexical distance	.69	.15	.48	.77

Note: All correlations are based on estimates for 62 languages except for morphological distance, which was available for 54 languages due to missing typological data.

Table 5 also includes correlations with lexical and morphological distances. With respect to the last two rows, lexical and morphological distance correlated more strongly with L2 learnability than both phonological similarity measures. The observation that both measures are more strongly correlated with each other ($r = .77$) than to new sounds or new features indicates that phonological similarity can potentially explain additional variance in L2 learnability on top of lexical and morphological distance measures. However, the new features measure also correlates with morphological distance and lexical distances ($r = .49$ and $.48$ respectively).

Model comparison showed that the effect of phonological distance on L2 learnability persists after we added lexical and morphological distance. The number of new sounds significantly adds to a model that already contains lexical distance ($\chi2$ $(1) = 5.8$, $p < .05$), or both lexical distance and morphological distance ($\chi2(1) = 10.5$, $p < .01$). New features did not significantly add to a model that already contains lexical distance ($\chi2$ $(1) = 2.58$, $p = .11$), or both lexical distance and morphological distance $\chi2(1) = 3.76$, $p = .52$). Thus, although new features were a better measure than new sounds in isolation from lexical and morphological distances, new sounds were a better measure in conjunction with lexical and morphological distances.

Discussion

This study investigated the relation between L2 learnability of Dutch and phonological distance between the sound inventories of Dutch and 62 L1s. We used L2 proficiency scores controlled for age, exposure, gender, and education and aggregated across L1s as measures of L2 learnability. Our main finding was that typological differences between the sound inventories of the specific L1-L2 combinations involved successfully explain variation in L2 learnability. A lower number of new sounds and a lower number of new features point to additional benefits for L2 learners ($r = -.47$ vs. $r = -.35$). Following our two hypotheses, we conclude that the similarities between sound inventories facilitate L2 learnability. We discuss the importance of these findings for phonological accounts of L2 learnability, distance-based models of L2 learnability, and their relation to lexical and morphological distance measures.

Learning Distant Sounds

Phonological distances successfully index L2 learnability. The effect of the new features distance aligns with the hypothesized effects of underspecification between sound inventories that contrastive analysis had already identified (Haugen, 1966; Ternes, 1976; Weinreich, 1963). L2 learnability is lower the more new features L2 sounds have (underspecification effects). This suggests that similar new sounds pose lower learning requirements than distant new sounds.

Our findings are consistent with theories that assume that learners keep track of feature geometries of speech sounds. L2 learners need to expand their feature geometries to get access to new features of new sounds in order to correctly perceive and produce sounds with new features (C. Brown, 1998, 2000). This way, the sounds of the L1 constrain the difficulty of perception and production of L2 sounds. This process may explain what happens after learners fully transfer their L1 sounds to the initial state of L2 learning (Schwartz & Sprouse, 1996), and how learners remain susceptible to new sounds (Hancin-Bhatt, 1994). After the initial stages of L2 learning have passed and learners reach upper-intermediate levels of proficiency, the L1s of the L2 learners continues to influence generalization to new sounds.

The role of distance to new sounds may be related to the relative increase and decrease in size of the L1 and L2 acoustic space. For example, a larger L1 than L2 vowel inventory may be beneficial because of its extended acoustic space (Iverson & Evans, 2009). Generally, these findings fit to a framework of inference, as more variation in previously acquired languages facilitates generalization to a new language or linguistic variant (Pajak, Fine, Kleinschmidt, & Jaeger, 2014). This means that learners are able to use variation as a means to facilitate representation of new input. For an introduction to these ideas in general see Jaeger & Tily (2011). Accordingly, we predict that learners who want to optimize their learning process should focus on learning the different sounds, as these seem to impede L2 learnability the most. This way, learners with an L1 for which an L2 has many distant new sounds experience the highest learning challenge, as they need to spend more effort.

Distance-based Models of L2 Learnability

The role of phonological differences can now be studied with currently available data of the sound inventories and their feature representations of many languages at the same time. Our model assumed that the role of multiple competitors in learning new sounds is captured by taking the minimally distant sound. However, it might help to have multiple similar categories to disambiguate a new sound. If the interference comes from being pulled to similar categories, being pulled in multiple directions might help to establish a new category.

Furthermore, our measure of new features does not distinguish between features that are new for multiple sounds. For example, [labiodental] can be a new feature for both [f] and [v]. Accordingly, the number of additional contrastive features may play a role as well (Pajak, 2012). For example, when the feature [long] is present in the L1 on vowels only and the learner needs to acquire a long consonant, generalization from long vowels to long consonants may be relatively easier (Tsukada, Hirata, & Roengpitya, 2014). Our initial investigations suggest that learners who do not have the contrastive feature [labiodental] score significantly lower on Dutch speaking proficiency ($r = .51$, $p < .001$). However, most languages use most features contrastively for some minimal pair in their sound inventory.

Another assumption that we made was that Dutch sounds are all equally frequent although we know that their usage differs (Luyckx et al., 2007). Surprisingly, the sounds that are most often different across L1s and Dutch are also of relatively low frequency in Dutch (e.g. diphthongs rank consistently low on frequency of use and /ɣ/ ranks 26 in the Dutch sound frequency ranking of Luyckx et al., 2007). Less common sounds may be less important for a speaker's production of intelligible speech. Future work should also focus on the role of frequency, both in the L1 and in the L2.

In relation to frequency differences, the features of frequently occurring neighboring sounds may co-determine what sound combinations will be surprising to learners and which are to be expected. Future work may test additional effects of differences between phonotactic rules and syllable structures on L2 learnability as the suprasegmental L1 structures may also influence L2 learnability. For example, when adult speakers of L1 Dutch need to segment artificial speech, they benefit from native OCP-Place speech segmentation constraints, i.e. the constraint that consonants with shared place do not come in pairs (Boll-Avetisyan & Kager, 2013). Native speakers of Mandarin Chinese, which is not restricted by OCP-Place, do not make use of these constraints when they start to learn Dutch. The potential benefit from OCP-Place as a cue for L2 speech segmentation may depend on the prevalence of OCP-Place constraints in the language background of the learner (Boll-Avetisyan, 2012). Although, to the best of our knowledge, similarity between syntactic constraints has not yet been studied, we expect OCP-Place constraints to affect L2 learning.

Phonological, Lexical, and Morphological Distance

After lexical and morphological distance have been accounted for, a measure of new sounds still provides significant improvement of model fit whereas a measure of new features does not. Lexical and morphological distances correlate more strongly with L2 learnability than phonological distances.

Does phonological similarity specifically or overall similarity explain variation in L2 speaking proficiency? This issue bears on the question of the domain-specificity of the learning mechanism. Arguments against domain-specificity include its implausible evolutionary pathway (Chater & Christiansen, 2010) and that learners can achieve domain-specificity by starting from a set of domain-general learning mechanisms (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Understanding the domain-specificity of L1 transfer to an L2 is a focus in much of the research on transfer as it is challenging to disambiguate between different transfer effects on L2 learnability. Here, we have presented distance-based measures of phonological transfer, while other studies make lexical and morphological distance-based measures available (Schepens et al., 2013a, 2013b). Lexical and morphological distance correlate with the new features distance measure of phonological distance. We cannot provide definite answers as to which of these distances presents the greatest challenge to L2 learners: they are all successful indexes of L2 learnability. Due to effects of collinearity, we cannot be sure as to whether phonological distance effects are mediated by morphological similarity as well. However, this collinearity between different distance measures does not change the finding that new features successfully explain variation in L2 learnability, and are directly related to adult learnability of additional languages. In order to disentangle phonological and morphological effects, domain-specific evaluation of pronunciation and grammatical proficiency may be helpful.

Conclusions

The evidence that we present comes from a comparison of 62 languages using data from language-testing and phonological typology. The reported measures of phonological distance are useful for comparison between languages from different families. Our approach allowed the evaluation of phonological distance measures at the level of sound inventories as well as the level of features. The phonological distance measures are distance-based models of the role the L1 plays in learning L2 sounds. We argue that distance plays an essential role in the transfer process underlying the learnability of an additional language. Our main conclusion is that learners use information at the level of the phonological feature inventory to generalize to the new sounds of an L2.

In addition, we conclude that the heterogeneity of state exam data can be used as an advantage to generalize over individual differences. State exam data nicely complements experimental approaches to language background effects involved in L2 learnability. Typological data provide the opportunity to investigate the ways of accounting for variation in L2 learnability across many L1s simultaneously.

Chapter 5

The L2 impact on Acquiring Dutch as an L3: The L2 Distance Effect

Acknowledgments
We thank the audience of the Leuven Statistics Days (June 2012).

This chapter was submitted as:
Schepens, J., Van der Slik, F., & Van Hout, R. (submitted). The L2 impact on acquiring Dutch as an L3: the L2 distance effect. Speelman, D., Heylen, K. & Geeraerts, D, (eds). *Mixed Effects Regression Models in Linguistics*. Springer: Quantitative Methods in the Humanities and Social Sciences.

Abstract

Cross-classified random effect models (CCREMs) are often used for partitioning variation in experimentally collected as well as in cross-sectional linguistic data. However, crossed random effects may have complex interrelationships, more complex than is usually assumed. This becomes clear in comparing L1 and L2 influences on proficiency in Dutch as an L3. Using a large database of L3 proficiency scores, we assessed the mutual dependency between the crossed random effects of L1 and L2. The results suggest independent and robust linguistic L1 and L2 distance effects: the smaller the linguistic distance, the higher the proficiency, and the L2 effect being weaker than the L1 effect. A model that incorporates an additional L1 by L2 random interaction effect still stipulates the relative importance of the L2 distance effect. The L1 distance effect is robust against the L2 distance effect and the L2 distance effect is robust against interactive effects. We discuss possible explanations for interactions between the L1 and the L2. The data support independent linguistic distance effects of both the L1 and L2, besides L1-L2 interactions. We recommend researchers to compare the fit of their crossed random effects models to the fit of models that include their interaction effects as well.

The L2 impact on acquiring Dutch as an L3: The L2 distance effect

Introduction

Cross-classified random effect models (CCREMs; Raudenbush & Bryk, 2002) are becoming the standard for analysing linguistic data (Barr et al., 2013). A paper that introduced CCREMs for psycholinguistic study under the header mixed effects models (Baayen, Davidson, & Bates, 2008) had been cited 1923 times already (as reported by Google Scholar September 2014). CCREMs have been used in studies of linguistic variation (Wieling, Nerbonne, & Baayen, 2011), syntactic variation in language production (Jaeger, 2008), cognate effects in bilingual word recognition (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010), etc. Few studies have paid attention to possible interrelatedness between random effects, with the exception of some school effectiveness studies (Leckie, 2009; Raudenbush & Bryk, 2002; Shi, Leite, & Algina, 2010; Van Tubergen & Kalmijn, 2005). The consequences of interrelated random effects are studied here by modeling the effectiveness of different language backgrounds (L1-L2 combinations) on proficiency in Dutch as an L3, similar to school effectiveness studies (e.g. Goldstein, Burgess, & McConnell, 2007).

Being able to use an additional language is widely regarded as helpful for successful integration and economic mobility in a foreign language environment. About half of the citizens in the EU Member States are able to hold a conversation in at least one additional language according to three reports carried out for the EU (47% in EC, 2001, 56% in 2006, 54% in 2012). Additional language use could benefit from a better understanding of the degree to which both the L1 and the L2 have an effect on acquiring proficiency in an L3. The question of the roles of L1 and L2 in L3 perception and production (L1-L2 interrelatedness) has often been addressed in psycholinguistic experiments, which are typically characterized by low numbers of participants and low numbers of L1s and L2s (Bardel & Falk, 2007; Cenoz et al., 2001; R. DeKeyser, 2012; Oldenkamp, 2013; Ringbom, 2007). In addition, participants are often asked to self-report their level of proficiency in an additional language. One way to overcome limitations of low numbers of participants and subjective language proficiency judgments, is to use formalized language testing proficiency scores when they are available for large numbers of subjects (Finnie & Meng, 2005). Until recently, only a few studies have made use of state databases for higher education, most often for automatic error detection (Nicholls, 2003; Yannakoudakis, Briscoe, & Medlock, 2011), but also for assessing the effect of linguistic distance (Van der Slik, 2010). With respect to L1-L2 interrelatedness, it has been found that the L3 is jointly influenced by a naturally acquired L1 and an educationally acquired L2 (Bardel & Falk, 2007). However, L1-L2 interrelatedness has not been investigated at a large scale using CCREMs. The primary aim of this study is to extend our knowledge of how L1s and L2s affect proficiency in Dutch as an L3. However, since the structure of this problem is similar to other problems in linguistics and other areas, the present study may also be useful for researchers currently using CCREMs for modeling behavioral data with a complex structure of interrelated random effects.

The degree to which language background influences acquisition of an additional language (cross-linguistic influence, CLI) has been “wreathed in controversy” since the emergence of second language acquisition research (Kellerman, 1995, p. 125). Currently, large databases are becoming available from language testing institutions, providing hitherto unprecedented opportunities to study complex interrelationships between L1s, L2s, and an L3. In a previous study in which we used CCREMs (Schepens et al., 2013b), we demonstrated that linguistic distance between the L1 and Dutch has a substantial and systematic impact on proficiency in Dutch by controlling for third factors related to the individual learner and characteristics of the country of origin. Linguistic distance can be modeled using measures of evolutionary change (Bouckaert et al., 2012; Holman et al., 2011) for testing whether an L1 distance effect explains differences in acquiring Dutch as an additional language. The impact of other background languages remained untested. We extend our model of proficiency in Dutch as an additional language (L3) by testing if the best additional language (L2) has an effect of its own and if such an effect can be explained by an effect of linguistic distance, in addition to the patterns previously observed across mother tongues (L1s).

Background

CCREMs with Interrelated Random Effects

Cross-classified random effect models (CCREMs) are multilevel regression models with crossed random effects that are not completely contained within one another (Beretvas, 2011; Bolker et al., 2009). For example, English is a common second language (L2) for native speakers of such languages as German and Spanish. However, native speakers of these languages speak other second languages besides and instead of English as well. When we investigate L3 Dutch proficiency scores, we can assume (as we do in this study) that the L1s and the L2s involved are drawn randomly from a larger set of languages, preferably from the distribution of all languages around the world. Hence, we call them random effects. Random effects are factors of which the levels are not fixed but randomly sampled. Consequently, if we treat both the L1s and the L2s of the candidates as independently crossed random effects, we assume that the variation in speaking proficiency across L1s and L2s follow independent normal distributions. However, as will be seen in this chapter, real world data sets are often more complicated. One under-investigated issue is that the variation across the levels of a random effect can depend on the variation across the levels of another random effect. For example, although L2 English is common for both L1 German and L1 Spanish, the variation in L3 Dutch proficiency scores due to L2 English is not constant across L1s (e.g. L1 Spanish learners may benefit more from L2 English than L1 German learners do). It is clear that the impact of a second language

(L2) on learning a third language (L3) cannot be studied without taking into account the impact of the first language (L1) of a language learner.

In CCREMs, random effects are assumed to vary independently. Each of the random effects has unknown variance components. A posteriori, these variance components can be estimated, and they are usually called best linear unbiased predictors (BLUPs; G. Robinson, 1991). In a CCREM, a response score is predicted by fixed coefficients, level 2 random effects, and level 1 residuals (Leckie, 2009; Rasbash & Browne, 2008). Often, CCREMs assume independence between random effects, sometimes due to a lack of data (Raudenbush & Bryk, 2002), sometimes by experimental design (Baayen et al., 2008). However, the consequences of assuming that random effects are completely mutually independent are not well understood (Meyers & Beretvas, 2006; Shi et al., 2010). One way of investigating interrelatedness between random effects is to incorporate an x-by-y random interaction effect, where x and y are the crossed random effects (Raudenbush & Bryk, 2002). However, a sufficient amount of data may not be available to reliably estimate every x-by-y combination in which the random effects are only "partially balanced" (Pinheiro & Bates, 2000), or "partially crossed" (Raudenbush, 1993). Consequently, many studies avoid taking into account x-by-y random interaction effects.

Interrelated L1 and L2 Effects

We want to contrast evidence for an independent L2 effect to evidence in favor of an interactive L1-L2 effect where the L2 depends on the L1 via L1-by-L2 random interaction. There are reasons to predict differences between an L1 and an L2 effect. Most importantly, L2 learning problems are more commonplace than L1 learning problems: Although virtually all adult L1 learners reach native proficiency levels, many L2 learners even struggle to learn train station basics in a foreign tongue. First, we hypothesize that the L2 plays a role beyond the L1. Having command of a second language is considered beneficial for further successive language learning. Secondly, we hypothesize that L1 and L2 effects are generally additive in contrast to an interactive account, meaning that the combination of acquired languages explains L3 proficiency better than either the L1 or the L2. Among the possible explanations various specific interaction patterns may occur. Third, we hypothesize that the importance of the L2 is less than that of L1. Fourth, we hypothesize that the effect of the L2 takes the form of a distance effect (Cenoz et al., 2001; Ringbom, 2007), similar to the L1 distance effect (Schepens et al., 2013b). In all, we have formulated four questions that we seek to answer, given the differences in L3 Dutch speaking proficiency scores across L2s

1. Is there an L2 effect on L3 Dutch speaking proficiency scores?
2. If so,
    a. Is the L2 distance effect better characterized as an additive, independent effect or is it an effect that needs to be explained in combination with the L1s involved?
    b. Is the L2 distance effect more or less important than the L1 distance effect for explaining differences in proficiency in L3 Dutch?
    c. Does the L2 effect follow a pattern that is consistent with linguistic distance?

Methods

We fit CCREMs on speaking proficiency test scores from the Dutch state exam “Dutch as a Second Language”. This state exam is developed by the Central Institute for Test Development (Cito) and the Bureau of Inter-Cultural Evaluation (Bureau ICE), two large test battery constructors in The Netherlands. The state exam is required for immigrants who want to enroll in a Dutch university, but the exam is also taken by many immigrants who come to the Netherlands for work or marriage. The speaking part of the exam is passed when the requirements for the B2 level on the Common European Framework of Reference (Council of Europe, 2001) are met (Kuijper, Bergsma, &

Bechger, 2004). This is comparable to an International English Testing System (IELTS) score of 5.5.

The speaking exam consists of 14 tasks that have to be completed in 30 minutes. The participants have to provide information, give instructions and so on. Professional examiners evaluate the speech on content and correctness according to formalized testing criteria. The participants can voluntarily fill in a brief questionnaire in which they are asked about various background characteristics. We use the responses on the questions about length of their residence in The Netherlands, age of arrival, gender, years of full-time education, country of birth, mother tongue, and best additional language. Best additional language represents the answer to the question: “If you speak another language besides your mother tongue, which other language do you speak? If you speak more than one other language, name the language that you know best.”

We use a sample of L3 speaking proficiency scores that was collected over the years 1995 to 2010 inclusive. We selected the first speaking proficiency score for 50,500 unique participants, as some participants try the same exam multiple times. We included only L1s, L2s, and countries of birth with at least 15 available scores. This resulted in a selection with enough data to test learning differences across 73 L1s (median 204.0 speakers per L1), 43 L2s and monolingual (median 57.5 speakers per L2, including monolingual), and 122 countries of birth (median 128 speakers). Out of the 3212 possible L1-L2 combinations, 759 L1-L2 combinations were observed in the data (216 combinations had at least 15 participants), see Table 1 for the 10 most common L1-L2 combinations. 35.7% of all L1 speakers have an L2 other than the most common L2, illustrating the cross-classified nature of the data. Excluding English as an L2, the data becomes only slightly more cross-classified (39.2% speak a different L2 than the most common L2). Candidates with missing answers on the questionnaire were removed from the analysis. Candidates with outlying speaking proficiency scores were also removed. The speaking scores were normally distributed, see Figure 1.

Our previous study already showed that speakers of closely related languages obtain better scores than speakers of less closely related languages and that immigrants from European countries perform better on average than immigrants from other continents (Schepens et al., 2013b). For example, although both come from Switzerland, Swiss native speakers of German perform better on average than Swiss native speakers of French. Furthermore, Spanish native speakers of Spanish perform better on average than Peruvian native speakers of Spanish. Further exploratory analyses show that, in general, bilinguals outperform monolinguals. For example, Americans who speak L1 English and L2 German perform better on average than Americans who speak L1 English and no other language. Moroccans who speak French perform better on average than Moroccans who speak Arabic only. Russians, Iranians, Afghans (etc.) who speak English perform better on average than their monolingual counterparts do.

We performed CCREM analysis in order to investigate the differences across L1, L2, and L1-L2 combinations more generally. We first estimated a CCREM without incorporating the L2 at all, so L1s were crossed with countries of birth (C). This means that we used the model in Schepens et al. (2013b), but without the binary indicator of L2 presence. This model included the fixed effects of gender, age of arrival, length of residence, years of full-time education, educational quality in the country of birth based on secondary school enrollment rates (UNESCO, 2011), and an interaction between the latter two covariates. Subsequently, models with more complex random effect structures were fitted to the data and compared using likelihood ratio tests based on -2 logarithms of the likelihood (-2LL) under a χ^2 distribution, which can be interpreted as measures of model fit: the probability of observing the data given the maximum likelihood estimations for the model.

The random effect structures were as follows. First, speaking proficiency was tested for dependency between the mother tongue (L1) and any additional language (L2), which together constitutes the language background (L1-L2) of the learners. In this test, we assessed whether a model with independently crossed random effects of L1 and L2 fits better or worse than a model with one homogeneous random effect of the L1-L2 (first and second model). The model with independent effects assumes that the effect of the additional language is constant and irrespective of the mother tongue the learner speaks. The model with homogeneous groups assumes that the effect of the additional language is variable and fully intertwined with the mother tongue the learner speaks. For the third model, we replaced the crossed effect of the L1 and L2 with a crossed effect of the L1 and L1-L2, effectively allowing for an intertwined effect of additional language to cross with the L1. For the fourth model, we added the crossed effect of the L2 back into the model.

In summary, four models were tested: a model with a random effect of L1-L2, a model with crossed random effects for L1 and L2, a model with crossed random effects for L1 and L1-L2, and a model with crossed random effects for L1, L2, and L1-L2. To all of these four models, a crossed random effect of country (C) was added as well. All these models included the fixed effects described in the previous paragraph. The fixed effects were added in order to separate variance that is due to language background from confounding variables related to individual and country characteristics.

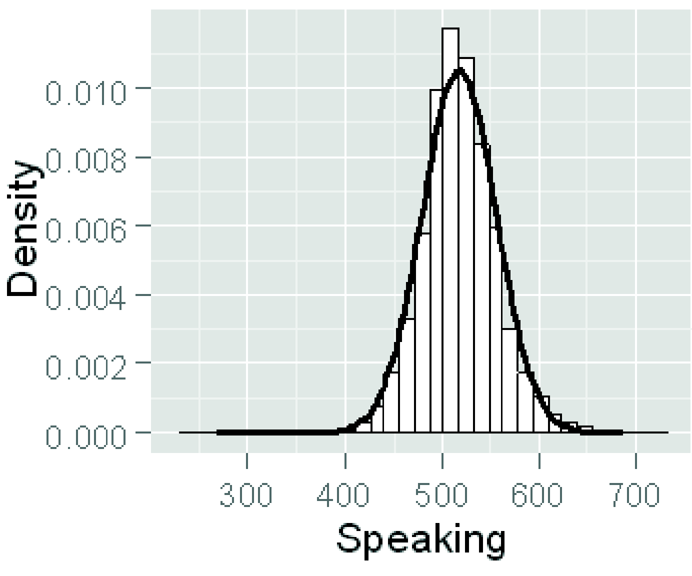


Figure 1. Distribution of speaking proficiency scores.

Table 1. The 10 most common L1-L2 or L1-monolingual combinations. The fourth column shows the number of speakers for a given combination. Monolingualism is present in the top 10 on positions 7 and 9.

Rank	L1	L2	Top N
1	German	English	4336
2	Arabic	French	2933
3	Russian	English	2439
4	Arabic	English	2036
5	Spanish	English	1976
6	Polish	English	1733
7	English	-	1666
8	Persian	English	1646
9	Turkish	-	1436
10	Serbian	English	1174

Results

Several CCREMs were fitted to the data using different random effect structures as described above. The parameters were estimated using the *lme4* package in R (D. Bates et al., 2011): `lmer(Speaking~1 + Gender + Age of Arrival + Length of Residence + Years of Full-time Education * Educational Quality + (1|Country) + (1|L1), data)`. Note that the interaction term includes the separate covariates. We attempted to improve the fit of this model by comparing different ways of modeling the random effect structure. The model that fitted best to the data included random effects for the L2 `(1|L2)`, and L1-by-L2 random interaction `(1|L1-L2)`. We present results from model comparison and addition of control variables. After isolating L2 variation, the L2 distance effect is then demonstrated using orderings of random adjustments and pairwise comparisons.

Model Comparison

Table 2 provides several measures of model fit used for modal comparison. Several CCREMs are compared to explore the effects of varying the random effect structure. First, three different null models were fitted in order to evaluate the effect of controlling for confounding variables, the hierarchical structure of the data, and the fit of a model without L2 related parameters, respectively. The overall null model (see null model 0-0 in Table 2) provides a baseline for the subsequent more complex models. The overall null model has four parameters: a fixed intercept, two random intercepts, and a parameter for residual (student level / level 1) variance. The other null models (null model 0-1 and null model 0-2) are baseline models to demonstrate the hierarchical nature of the dataset. In null model 0-1, the intra-class correlation is 27.5% (as computed using the values in Table 3). More complex random effect structures show that the hierarchical structure of the dataset is more complex due to by-L1 and by-L2 variation, leading to a ratio of 45.9% of (cross-classified) level-2 variation as compared to total variation in model 4. This demonstrates the importance of incorporating distinct classes in subsequent models, at the same time producing dramatic improvement as based on likelihood ratio tests. Null models 0-1 and 0-2 show highly significant improvements of fit as compared to the baseline null model 0-0 due to the added control variables (see Table 2). AIC (-2LL plus twice the number of parameters in the model) and BIC (-2LL plus the number of parameters times the natural logarithm of the sample size) are added for additional comparison. Controlling for six confounding variables (see Methods and introductory paragraph of Results) explains part of the variation across countries of birth (32.8%) and mother tongues (5.6%) as compared to the baseline null model 0-0. Since no measures of linguistic differences are added, it is not surprising that explained variance across mother tongues lags behind (Schepens et al., 2013b). Table 3 further shows the estimates of the variance components as well as 95% HPD intervals in parentheses, quantifying confidence in the shown parameter estimates. According to the criteria mentioned in various sources (Baayen, 2008; Goldstein et al., 2007; Leckie, 2009), the widths of the reported HPD intervals offer no reason to remove any of the parameters from any of the baseline or other reported models. As an additional check, all parameter estimates were inspected for bimodal patterns using density plots and for deviations from normality.

With the remaining four models (models 1 to 4 in Table 2 and 3, several ways of modeling L1-L2 interrelatedness are compared. The first model assumes that each language background is unique and interactive, and that it is not possible to identify by-L1 or by-L2 variance separately. The second model assumes the opposite, namely that it is not possible to identify an L1-by-L2 random interaction effect. It assumes that by-L1 and by-L2 variance is additive and independently contribute to L3 proficiency. The data provides significantly more support for the second model than for the first model. This result shows that there is an L2 component in the variance across L3 proficiency scores (question 1), and that there is more evidence for an additive effect than for an interactive effect (question 2a). The parameters also show an increase in the proportion of language-to-country variation, suggesting that the gain in model fit can be attributed to the allocation of remaining variance to a combination of by-L1 and by-L2 variance. The third model assumes that a random interaction effect is a better explanation than an L2 effect. This is confirmed by the data. Furthermore, the by-L2 adjustments depend on by-L1 adjustments. However, an even more complex model fits the data best: By allowing both random interaction and by-L2 adjustment, the fourth model assumes that an L1-independent L2 effect still plays a significant role, alongside random interaction effects. The increase in model fit is again highly significant, showing the importance of the L2, independently of the L1. The estimated parameters for the by-L1 and by-L2 variance indicate that a larger proportion of variance can be attributed to L1 factors than to L2 factors (question 2b). Next, after first describing the role of control variables, we will then assess the role of linguistic distance in the by-L1, by-L2, and L1-by-L2 variance (question 2c).

Table 2. Likelihood ratio tests showing significant improvements of fit against the chi-squared distribution, irrespective of the clearly increasing complexity in random effect structure of the CCREMs. AIC, and BIC information criteria are added for reference. All comparisons are significant at the .0001 level.

Model No.	Random Effects	df	AIC	BIC	-2LL	χ^2
0-0	L1, C	4	495,100	495,136	495,092	
0-1	C	9	493,456	493,536	493,438	1654.08
0-2	L1, C	10	492,692	492,781	492,672	765.90
1	L1-L2, C	10	492,214	492,302	492,194	478.33
2	L1,L2, C	11	492,030	492,127	492,008	186.43
3	L1,L1-L2, C	11	492,008	492,105	491,986	21.35
4	L1,L2,L1-L2, C	12	491,914	492,020	491,890	96.21

Table 3. Parameter estimates of the variance components for the intercept-only CCREMs demonstrating that variation across random effects depends on the total random effect structure incorporated in the model. The standard deviations are restricted maximum likelihood (REML) estimates and the widths of 95% highest posterior density (HPD) intervals (in parentheses) are constructed from model-specific chains of 20,000 Markov Chain Monte Carlo (MCMC) samples each.

No.	Residual	C	L1	L1-L2	L2
0-0	32.37 (0.40)	12.97 (4.21)	12.40 (4.21)		
0-1	31.90 (0.40)	12.09 (3.10)			
0-2	31.63 (0.39)	8.72 (2.63)	11.71 (3.89)		
1	31.35 (0.39)	8.84 (2.72)		10.84 (1.76)	
2	31.41 (0.39)	8.44 (2.68)	11.36 (3.86)		3.77 (2.45)
3	31.34 (0.39)	8.23 (2.63)	11.18 (3.99)	5.96 (1.48)	
4	31.34 (0.39)	8.30 (2.72)	11.13 (3.98)	3.29 (1.45)	3.82 (2.60)

Control Variables

Table 4 shows the estimated parameters for the fixed part of model 4. Included in the model are 12 parameters, including six fixed control variables and a fixed intercept. We incorporated the control variables to control the estimations of the random effects for individual differences. The control variables were not centered or otherwise normalized. All the effects were highly significant (apart from effect number 4, see Table 4). The gender effect represents a change of 7.39 points in the L3 proficiency score (see Figure 1 for the scale), associated with a participant being female, all other predictors being equal. See Table 4 for 95% HPD intervals. Also beneficial were an earlier age of arrival, a longer length of residence, a shorter full-time education (non-significant, see also Chapter 6), a higher educational quality, or a high combination of full-time education and quality. Although not relevant for the estimates of interest here, we report collinearity between the control variables 4 and 5 ($r = -.50$), 5 and 6 ($r = -.53$), and 4 and 6 ($r = -.95$), explaining why the sign of full-time education is in an unexpected direction. Models with random slopes were tested but these converged only sporadically. Explorations into various estimated variance and covariance structures revealed only small variations in the way the fixed effects estimates deviated across the random effects. The explorations suggested that age of arrival and the education-related effects varied and co-varied across the random effects.

Table 4. Parameter estimates for the fixed predictors included in model number 4. All estimates are significant at the .0001 level apart from full-time education. The HPD intervals were constructed from one 20,000-sample MCMC chain.

Fixed Effects	Estimate	2.5% HPD	97.5% HPD
0. Intercept (points)	505.02	498.44	511.36
1. Gender (1 = Female)	7.39	6.74	8.05
2. Age of Arrival (years)	-0.72	-0.77	-0.68
3. Length of Residence (years)	0.62	0.55	0.69
4. Full-time Education (years)	-0.77	-1.83	0.24
5. Educational Quality (% gross)	0.18	0.11	0.25
6. Interaction 4* 5	0.04	0.02	0.057

The L2 Distance Effect

We can now isolate the part of the variance in L3 Dutch proficiency scores that is due to differences across L2s. Figure 2 presents the contribution of the by-L2 adjustments to predicted proficiency scores. It shows how the model distributes the estimated L2 variance component of model 4 across the unique L2s. The predicted proficiency scores (dots) incorporate the by-L2 adjustments of model 4. The lengths of the black lines represent the relative benefit of speaking for each of the 18 depicted L2s. Only the 18 most frequently spoken L2s are shown. The benefit of speaking German is highest (+10.15, see also Table 6), the benefit of speaking no L2 is lowest (-5.20 points), Turkish as an L2 is second to lowest (-4.39). Figure 3 shows the relationship of by-L2 adjustments with by-L1 adjustments ($r = .60$, $p <$ 0001). The graph makes a number of interactions visible; in particular, German is highly beneficial as a second language and Italian is only of relatively low benefit for learning Dutch. The slope of the plotted regression line suggests that the L2 distance effect is about 1/6 of the size of the L1 distance effect.

L1 and L2 distance effects were examined further by comparing the by-L1 and by-L2 adjustments (BLUPs) for different random effect structures. In Table 5, the top 10 by-L1 adjustments are ordered from high to low for the null model that included only country and L1 in the random effect structure (model 0-2). The second and third columns show what happens to these estimates after L2 is brought into the model. It becomes clear that only slight modifications to the estimated by-L1 adjustments are predicted after accounting for L2 variance. Two languages in the top 10 switch two positions. The L1 benefit of Estonian (underlined in table 5) decreases, whereas the L1 benefit of English increases. It may be the case that the L2s of Estonians (e.g. Russian), when spoken as an L2 by other speakers of other L1s, produce lower L3 proficiency scores. The L2s of native English speakers (e.g. German), on the other hand, may produce relatively high proficiency scores across other L1s (underlined in table 5). The table shows that the L1 effect is stable across simple and complex models, in particular for the model that accounts for L2 effects. Furthermore, the higher ranks show a preference for linguistically less distant languages, as characterized by the five Germanic languages that are included in the top six. As Slovenian is not widely spoken outside Slovenia, it may be the case that its schooling quality has been underestimated by our predictors or that most Slovenians on average speak more than two additional languages besides their mother tongue (EC, 2012).

In column 1 of Table 6, the top 10 by-L2 adjustments from model 1 (L1 and L2 only) are shown. As in Table 5, we assess the stability of the L2 distance effect by comparing by-L2 adjustments for model 1 with by-L2 adjustments for model 4 (after L1-by-L2 random interactions are added). The ordering is again robust, this time against the addition of L1-by-L2 random interaction effects. The estimate of the adjustment for English changes the most: it decreases. The L1-by-L2 random interaction may have taken over some part of the adjustment for English, i.e. the variation is decomposed differently, suggesting that the L2 distance effect is not stable for English. One explanation may be that L2 English proficiency is relatively variable as compared to L3 proficiency in other languages. In all, Table 6 shows that the L2 effect is robust against interactional effects. Included in the top six L2s are four Germanic languages. Although less clear than the L1 distance effect, there seems to be a non-random ordering in the benefits of L2s that follows the ordering of an L2 distance effect to considerable extent.

Pairwise comparisons are used to further illustrate the L2 distance effect for specific language backgrounds. These pairwise comparisons show that, on average, bilinguals who speak a closely related language to Dutch as an additional language score higher on speaking proficiency in Dutch than bilinguals with the same mother tongue but who speak a less closely related language to Dutch as an additional language. The pairwise comparisons show that this pattern holds for many L1s, with only few exceptions.

The pairwise comparisons were performed using aggregated random effects to compare the total of the random variance attributed to each L1-L2 combination. Aggregated random effects were computed by subtracting the fixed predicted proficiency (based only on the fixed predictors) from the fitted scores. We could not use either by-L1 or by-L2 adjustments as these capture independent variation only. For a number of L2s given an L1, we computed each time whether a specific L2 provided improvement over another one using t-tests between the aggregated random effects at the level of the learners, i.e. computing means, standard deviations, and number of learners for every L1-L2 combination.

The pairwise comparisons showed that bilinguals with L2 English usually performed better than bilinguals with L2 Russian. Pairwise comparisons revealed this pattern in native speakers of Bulgarian (T = 57.09, $p < 0.001$), Polish (T = 114.26, $p < 0.001$), Lithuanian (T = 1.62, $p = 0.108$), Serbian (T = 62.03, $p < 0.001$), Pashto (T = 7.32, $p = 2.99$), Armenian (T = 30.43, $p < 0.001$), and the reverse pattern in Persian (T = -15.30, $p < 0.001$). The beneficial effect of English as compared to Russian may result from a larger linguistic (and cultural) distance from Russian to Dutch. In addition, bilinguals with L2 English often performed better than bilinguals with L2 French. This pattern was found in native speakers of Polish (T = 19.99, $p < 0.001$), Serbian (T = 15.81, $p < 0.001$), Russian (T = 7.03, $p = 0.001$), and Spanish (T = 4.76, $p = 0.001$), although a reversed pattern was found for Portuguese (T = -2.71, $p < 0.01$) and German (T = -8.51, $p < 0.001$). In addition, L2 English was more beneficial than L2 Italian as is evident from the pattern in native speakers of German (T = 4.87, $p < 0.001$) and Spanish (T = 16.51, $p < 0.001$). L2 English was also more beneficial than L2 Spanish, as suggested by the pattern in native speakers of German (T = 5.86, $p < 0.001$), French (T = 6.44, $p < 0.001$), and Portuguese (T = 7.78, $p < 0.001$). Bilinguals with L2 German performed even better than bilinguals with L2 English. Pairwise comparisons were significant for native speakers of Czech, French, Polish, Slovak, Russian, Serbian, and Spanish. In addition, bilinguals with L2 German performed better than bilinguals with L2 French, as suggested by the pattern in native speakers of English and Spanish. Table 7 illustrates the rank ordering in L2s for five L1s mostly in line with predicted L2 distance effects. For example, for Serbian, L2 German is significantly more beneficial than L2 English, which in turn is significantly more beneficial than L2 French. L2 French is significantly more beneficial than L2 Russian, and no L2 (monolingual) is the least beneficial. The displayed means in Table 7 are all aggregated random effects, which can only be interpreted in comparison to the overall average of the random effects (i.e. L1 Serbian L2 German is 10.21 more beneficial than the overall average adjustment to the fixed predicted score). In all, bilinguals performed better than monolinguals in 29 out of 33 pairwise comparisons. 45 out of 50 pairwise comparisons, with on average 5 comparisons per language, showed a pattern that was consistent with our hypothesis that distance from Dutch to the additional languages determines proficiency in Dutch. The pattern was absent in some L1-L2 combinations, supporting L1-by L2 random intercepts.

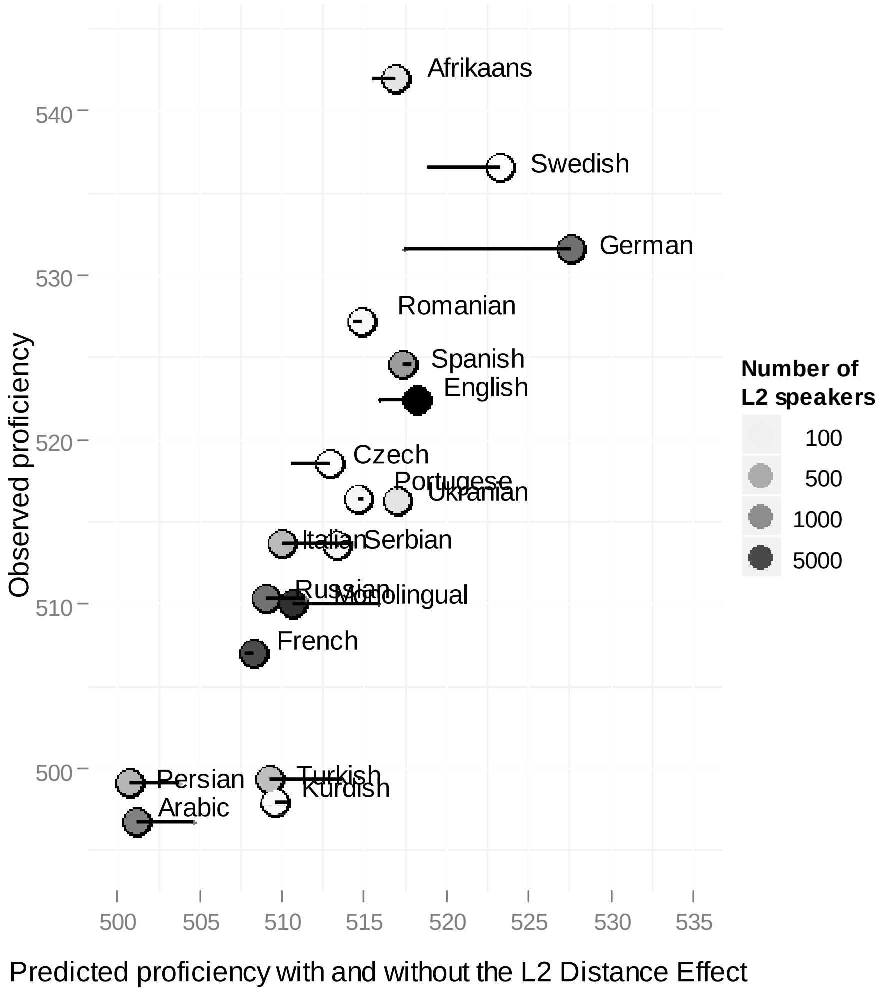


Figure 2. The by-L2 adjustments shift the predicted proficiency scores towards observed proficiency scores. The dots represent predicted proficiency of model 4 for L2 speakers of the labelled languages. The half tones represent the numbers of L2 speakers. The black lines represent the change in predicted (fitted) proficiency between model 4 (including by-L2 adjustments and control variables) and fixed predicted proficiency (including control variables only). Only languages (N = 18) with more than 70 L2 speakers are shown.

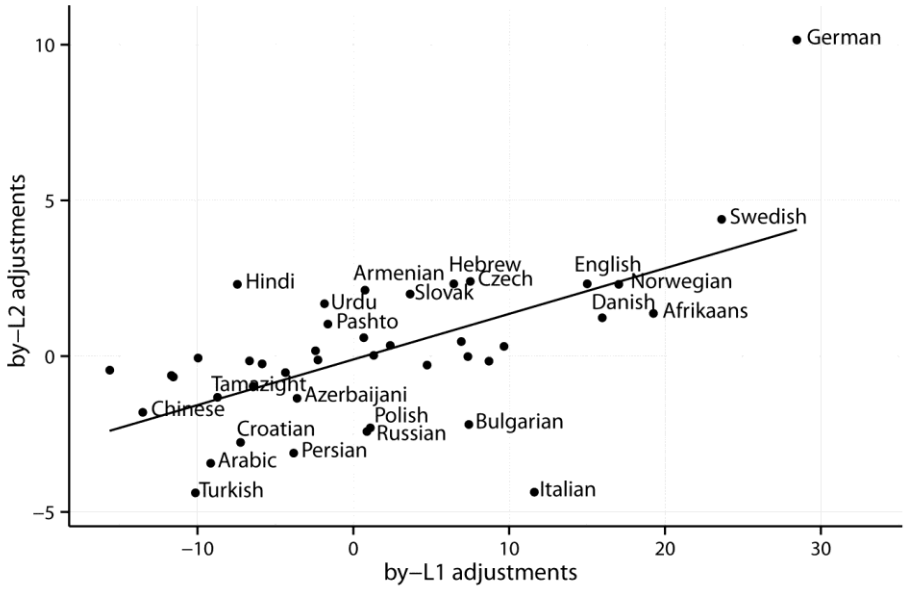


Figure 3. Scatter plot showing the relation between by-L1 and by-L2 adjustments (r = .60, p < 0001). Only L1s of which we have an estimate for the L2 are shown (using model 4). Text labels are shown for L2s with a higher or lower by-L2 adjustment than 1 point. The linear regression parameters are: y = .15x + -.11.

Table 5. The L1 distance effect is robust against incorporating L2 variance. The numbers are aggregated random effects taken from null model 0-2 (L1 only) and model 2 (L2 added). The largest positive and negative differences in the table are underlined.

Language	L1 only	L2 added	Difference
German	25.93	27.07	1.14
Swedish	24.97	24.24	-0.73
Slovenian	21.67	19.97	-1.70
Afrikaans	19.27	19.09	-0.18
Danish	18.96	17.56	-1.40
Norwegian	18.90	17.20	-1.70
<u>Estonian</u>	16.53	14.69	-1.84
Papiamentu	15.08	15.14	0.06
<u>English</u>	12.90	16.29	3.39
Belarusian	12.84	11.55	-1.29

Table 6. The L2 distance effect is robust against incorporating L1-by-L2 random interactions. The numbers are aggregated random effects taken from model 1 (L1 and L2 only) and model 4 (Interactions added).

Language	L1 and L2 only	Interactions added	Difference
German	10.11	10.15	0.04
Swedish	4.49	4.39	-0.10
English	3.33	2.32	-1.01
Czech	2.52	2.40	-0.12
Hindi	2.50	2.30	-0.20
Norwegian	2.46	2.30	-0.16
Hebrew	2.23	2.32	0.09
Slovak	1.94	2.00	0.06
Urdu	1.87	1.69	-0.18
Pashto	1.74	1.03	-0.71

Table 7. Pairwise comparisons (t-tests) of each estimated group adjustment with its immediately preceding estimated group adjustment (if the L1 is the same). The numbers are aggregated random effects taken from model 4 (Interactions added). The estimates are statistically controlled for educational differences. The L1s are displayed in no particular order.

L1	L2	estimation, p value
Kurdish	English	-1.77
Kurdish	Arabic	-7.92, $p < .0001$
Kurdish	Monolingual	-9.28, $p < .0001$
Kurdish	Farsi	-13.47, $p < .0001$
Kurdish	Turkish	-19.9, $p < .0001$
Serbian	German	10.21
Serbian	English	2.89, $p < .0001$
Serbian	French	-1.14, $p < .0001$
Serbian	Russian	-4.62, $p < .0001$
Serbian	Monolingual	-7.89, $p < .0001$
Hungarian	German	19.64
Hungarian	Romanian	18.79, $p < .0001$
Hungarian	English	16.93, $p < .0001$
Hungarian	Monolingual	4.73, $p < .0001$
Polish	German	9.44
Polish	English	5.06, $p < .0001$
Polish	French	2.53, $p < .0001$
Polish	Russian	-.85, $p < .0001$
Polish	Monolingual	-1.88, $p < .0001$
Polish	Italian	-3.59, $p = 0.008$
German	French	36.44
German	English	34.12, $p < .0001$
German	Italian	31.20, $p < .0001$
German	Spanish	31.06, $p = 0.90$
German	Russian	27.67, $p < .0001$
German	Monolingual	26.66, $p = 0.11$

Discussion and Conclusion

By varying the random effect structure of cross-classified random effect models fitted on a large number of language learners, we investigated the interrelatedness between the L1 and the L2 for learning L3 Dutch. The predicted proficiency scores indicate that a significant part of the variation is decomposed into independent L1 and L2 distance effects. Also, a significant part of the variation is further decomposed into an L1-by-L2 random interaction effect. However, comparing by-L2 adjustments for different models shows that the L2 distance effect is robust against interactive models. In addition, pairwise comparisons show that the L2 distance effect is observed repeatedly within different L1s. The findings are discussed below in terms of the concept of linguistic distance and additional language processing.

By-L1 and by-L2 adjustments are akin to each other ($r = .60$, $p < .0001$). As we have shown in a previous study (Schepens et al., 2013b), linguistic distance between the L1 and Dutch plays a decisive role in learning Dutch as an additional language (75.1% of explained variance). We now also know that, besides this L1 distance effect, L2 distance also has an effect on learning Dutch as an additional language (answering question 1 from the introduction), although this effect is about six times less strong (question 2b) as shown in Figure 3. The part of the variance modeled by random L1-by-L2 interactions suggests that L1-by-L2 random interactions still play a role. However, a model with independent L1 and L2 components fitted significantly better than an interactive model (question 2a).

Linguistic distance can be measured in terms of proficiency scores in a non-native language. We produced orderings of by-L1 adjustments as well as by-L2 adjustments that can be used successfully as an empirically validated measure of linguistic distance. Linguistic distance based on additional language proficiency is important for cross-linguistic influence studies (Cenoz et al., 2001; Ringbom, 2007) and for immigrant studies (Chiswick & Miller, 2005; Van Tubergen & Kalmijn, 2005). Linguists now have the opportunity to take into account empirically based measures of linguistic distance besides language classifications into families and genera or phylogenetic distances modeling the degree of evolutionary change between languages (Bouckaert et al., 2012; Holman et al., 2011).

Linguistically, the finding that the benefit of L1s and L2s on learning L3 Dutch is not constant introduces novel questions for further empirical research that may have consequences for the way we understand native and additional language processing. For example, the question arises whether some languages are better suited for non-native language processing than others. Additionally, as the data argues for an additive explanation, it seems that the individual types of languages in the mind determine learning more than the combination of types.

In all, we argue that incorporating interactions between random effects into CCREMs challenges the normally assumed independence between the different components of a random effect structure. Here, an unbalanced cross-sectional dataset produces considerable support for random interaction effects, which requires explanations otherwise not considered. For example, in the present case, we may have been confronted with differences in the stability of proficiency between L1s and L2s. This option is discussed further in Chapter 6. Interrelated random effects pose challenges to researchers analyzing data with a complex hierarchical structure that have consequences for the interpretation of parameter estimates.

Chapter 6

L1 and L2 Distance Effects in Learning L3 Dutch

Acknowledgements
This chapter greatly benefitted from comments by two anonymous Reviewers, Jared Linck, and several members of the HLP / Jaeger lab.

This chapter was resubmitted to *Language Learning* as:
Schepens, J., Van der Slik, F., & Van Hout, R. (resubmitted). L1 and L2 distance effects in Learning L3 Dutch.

Abstract

Many people speak more than two languages. How do languages acquired earlier affect the learnability of an additional language? We show that linguistic distances between the languages involved play a role. Both larger L1 and L2 distances correlate with lower degrees of L3 learnability. The evidence comes from a large number of speaking proficiency test scores on the state exam *Dutch as a Second Language.* The candidates speak a diverse set of first languages (L1s) and additional languages (L2s). Lexical and morphological distances explain 47.7% of the variation across L1s and 32.4% across L2s in multilingual learners. Cross-linguistic differences require language learners to bridge varying linguistic gaps between their first and second language competences and the target language depending on linguistic distance.

L1 and L2 Distance Effects in Learning L3 Dutch

Introduction

Besides factors such as age and exposure, learning an additional language appears to be more difficult the larger the linguistic gap between the first language (L1) and the target language (L2) is. The degree to which the L1 facilitates or impedes learning a specific additional language can be defined as L2 learnability. More generally, learnability of an additional language characterizes the learning difficulty of a target language, depending on previously learned languages. It was found earlier that variation in L2 learnability across individual learners depends on the linguistic distance between their L1s and the target language, while controlling for effects of age, exposure, education, and gender (Schepens et al., 2013a, 2013b; Van der Slik, 2010). Linguistic distance can be defined as a measure that quantifies how distinct linguistic structures are, e.g. at the lexical or at the morphological level.

The concept of L2 learnability may be applied to learning or acquiring any additional language, be it an L2, L3, etc. Stated as such, however, this is an oversimplification of how the different language configurations play a role when additional languages are acquired. A growing number of studies in the field of L3 learning provide evidence for the impact of both the L1 and an additional L2 on acquiring another additional language, the L3. Accordingly, the relative impact of the L1 as compared to the L2 may depend on: 1) which language is more similar to the L3 (Ahukanna, Lund, & Gentile, 1981; Rothman, 2011; Singleton, 1987), 2) the degree of proficiency in the L1 and L2 (Lindqvist, 2010; Ringbom, 2007; S. Williams & Hammarberg, 1998), or 3) the status of the L2, as being relatively important by itself (Bardel & Falk, 2007; Flynn et al., 2004; Hammarberg, Cenoz, Hufeisen, & Jessner, 2001). Research focusing on the status of the L2 investigates whether a language influences L3 learning more when the language in question has been learned as an L2, compared to when it has been learned as an L1. For example, the effect of German on learning L3 Dutch may be stronger for speakers of L1 German than for speakers of L2 German, because prior L1 knowledge is relatively more important than prior L2 knowledge. Alternatively, it has also been proposed that L2 knowledge can block transfer of L1 knowledge (Bohnacker, 2006; Falk & Bardel, 2011).

In this contribution we want to investigate adult language acquisition of Dutch from the L3 perspective, taking into account both the L1 and, if present, the best spoken L2. We investigate L3 learnability in relation to the distances between the L1 and L2 on the one hand and the L3 on the other hand. We investigated three subsequently more specific hypotheses. Our **main** hypothesis is that the L2-L3 distance is a robust factor in explaining proficiency differences in learning a new language, the L3, in addition to and independent from the impact of the L1-L3 distance. Confirming this hypothesis implies that there is a robust effect of being multilingual as well, in the sense that multilinguals and monolinguals differ in performance. However, is being multilingual facilitative or impeding? In order to answer this question we need to compare the multilinguals to a monolingual baseline state. Our **second** hypothesis is that the L2-L3 distance effect is more facilitative the lower the distance is, as multilingual learners obtain higher target language proficiency scores in Dutch than monolinguals. An important question is how the distinction between mono- and multilinguals can be implemented in our statistical models. An obvious choice is to just make a distinction between mono- and multilinguals, but, in addition, we want to investigate the possibility of including the monolinguals in the L2-L3 distance measures. A substantial group of learners has no additional language to their native language, next to Dutch. These learners started to learn Dutch as monolinguals. We want to investigate the possibility of including the monolinguals in the analyses of the L2-L3 distance measures. We have to assign them some L2-L3 distance value. Furthermore, confirming the hypothesis that there is an L2 distance effect does not, in itself, speak to the issue of how strong this effect is compared with the L1 distance effect. Our **third** hypothesis states that the L2 distance effect is weaker than the L1 distance effect, as the L1 is generally learned earlier and more intensively. The hypotheses presented here address the question as to what extent multilingual learners benefit from general effects of being multilingual and to what extent the effect of L2 distance is facilitative as compared to the effect of the L1.

We used the same collection of speaking proficiency testing scores on the state exam *Dutch as a Second Language* of immigrants from virtually all over the world that were used by Schepens et al (2013a, 2013b) and we tested the explanatory power of both lexical and morphological linguistic distance measures. We investigated L3 learnability by controlling for confounding variables, both on the level of the individual learner (exposure, age, gender, education) and the country of origin of the learner (educational quality). Education and literacy play prominent roles in additional language learning. However, as reading and writing tests are part of the exam as well, we can safely assume that illiteracy is non-existent among candidates. The exam assesses Dutch speaking proficiency at the B2 level (Council of Europe, 2001) by evaluating if participants can produce adequate spoken language in a series of different communicative situations. We will refer to the state exam data as STEX throughout this chapter. STEX is a unique resource for estimating the factors involved in speaking proficiency as it is large-scale, provides details on individual and contextual characteristics of the L3 learners, and the L3 proficiency scores are official language testing scores, which are more reliable than self-reported proficiency scores. As is usual for language tests, the outcome measure is a general measure based on assessments of morpho-syntactical and phonological form, content, and vocabulary. STEX includes speaking proficiency scores for more than 50,000 learners of Dutch, collected over a period of 15 years. During this time, learners could answer questions about their language background on a voluntary basis. Candidates reported their mother tongue and their best additional language, in case they spoke another language besides their mother tongue and Dutch. STEX is discussed in more detail in the Background and Methods sections.

STEX enables us to test predictions related to current L3 acquisition theories in a detailed way. We will first test to what extent L1 and L2 linguistic distance measures can provide an accurate explanation of L1 and L2 effects in L3 learnability. We start by assuming that the L1 and L2 effects are additive and we proceed by testing more complex interactive models that give room to a distinctive role of specific L1-L2 configurations. The additive model implies that an L2 may bring the learner linguistically closer to L3 Dutch, as an independent effect. For example, the distance between monolingual Spanish and Dutch may be larger than the distance between bilingual Spanish – French and Dutch, as French brings the learner linguistically closer to Dutch. This study operationalizes linguistic distance (typological similarity) by making use of comparisons of lexical distance and morphological complexity between the L1 and the L3 and the L2 and the L3. In line with current theories of L1 and L2 influence, we test the relative importance of the L1 versus the L2 (Flynn et al., 2004), which aligns with all three of our hypotheses (L2 is a robust factor, L2 is facilitative, L1 is more important). In addition, we test whether only the closest of the two determines L3 learnability (Rothman, 2011), which contrasts with our main hypothesis, and we test whether the L2 blocks L1 transfer (Bardel & Falk, 2007; Bohnacker, 2006), which aligns with a variation of our main hypothesis. This study contributes to understanding how L1 and L2 interact in L3 learnability, i.e. cross-linguistic influences in third language acquisition (Cenoz et al., 2001), and tests L1 and L2 effects as predicted in L3 learning theories. Our comparative approach is different from longitudinal studies of L3 learnability in adults. Rather than comparing target language proficiency over time, we compare target language proficiency across language backgrounds.

In the next section, we discuss current issues in L3 learnability, linguistic distance, and large-scale studies of speaking proficiency. This is followed by a description of the methods. We then present the statistical analyses of L3 Dutch state examination data. The concluding sections comprise a general discussion and the conclusion.

Background

L3 Learnability

Does an L2 have an influence on speaking proficiency in an L3 and, if so, is its effect different from L1 influence? The L3 literature suggests that both L1 and L2 typology in relation to the L3 and the L1 and L2 proficiency levels play a role (Cenoz, 2003; Jaensch, 2013; Murphy, 2005). It is unclear whether L2 typological similarity is more or less important than L1 typological similarity, and whether L2 proficiency is more or less important than L1 proficiency. In addition, L2 influence may differ between productive and receptive modalities, between written and spoken language use, and across learning stages. Learner-based variables, e.g., motivation, intelligence, years of full-time education, educational quality, age, gender, play a cardinal role as well. The present study focuses on speaking proficiency and on language-related variables, thereby controlling for learner-based variables by means of statistical tools. On a contextual level, we will also control for systematic variation for learners' countries of birth.

The current understanding of typology effects on L3 learnability predicts that a typological similarity or overlap between languages leads to positive cross-linguistic influences, both for L1 to L2 influence (Ard & Homburg, 1983; Kellerman & Sharwood Smith, 1986; Odlin, 1989), and for L2 to L3 influence (Cenoz, 2001). L1 negative transfer is more likely at lower proficiency levels (Odlin, 1989). There is not always a one-to-one correspondence between objective typological similarity and the typological similarity perceived by learners. A negative perception of typological similarity may lead to negative transfer (Jarvis & Pavlenko, 2008), overall or in a specific linguistic domain. The present study focuses on objective typological similarity.

There are at least three specific explanations for the way typological similarity influences L2 to L3 transfer. These are primarily based on the acquisition of syntactical properties, i.e. negation placement (Bardel & Falk, 2007), verb second (Bohnacker, 2006), relative clauses (Flynn et al., 2004), and word order (Rothman, 2010). The cumulative enhancement model (Flynn et al., 2004), first of all, predicts that the effect of the L2 (1) is not absorbed by the L1, (2) is either neutral or positive, and (3) is more beneficial for learning an L3 than having no L2 at all. In addition, the L2 status factor model (Bardel & Falk, 2007) and the findings of Bohnacker (2006) predict that an L2 can obscure or impede transfer effects between the L1 and the L3, depending on L2 status and L2 proficiency. This suggests a prominent role for the L2, outranking L1 influence. The typological primacy model (Rothman, 2010), finally, predicts that either the L1 or the L2 will transfer, depending on the highest typological similarity (see Table 3 in the Discussion for an overview of these predictions). Besides L2 transfer of syntax, L2 transfer has also been observed in the lexical domain of nonnative function words (De Angelis, 2005). We will investigate the different models by testing the additive and/or interactive impact of the L1s and L2s on learning Dutch as an L3.

In addition to L1 and/or L2 effects, individual differences determine a large part of the variation in L3 proficiency. Age effects have received considerable attention due to a general interest in critical period effects in L1 and L2 learning. Contrary to the hypothesized critical period for L2 learning, the critical period seems to lead to only small proficiency differences in ultimate L2 language proficiency in observational data (Bleakley & Chin, 2010; Hakuta et al., 2003) and allows adult learners to attain native-like accents (Bongaerts, 1999; Bongaerts et al., 1997). In addition, it is widely accepted that language background affects additional language learning, although it is unclear how important these effects are in comparison to traditional factors of interest such as age at onset of acquisition and duration of exposure.

Does the mind structure L2 knowledge in a similar way as it structures L1 knowledge? This is a relevant question since typological similarity between the L1 and the L3 can be of lower importance than L2-L3 similarity for learning an L3 (Bardel & Falk, 2007). This L1-L2 difference in transfer is hypothesized to result from the representational nature of L2 knowledge: adults acquire L2 knowledge initially on an explicit / declarative basis, before they can acquire it implicitly (Ellis, 2005; M. Paradis, 2009; Ringbom, 2007), which may be more beneficial for L3 learning than procedural L1 knowledge. In addition to procedural L1 knowledge, if L1 or L2 declarative knowledge is available, it may be more likely to affect learning of the declarative knowledge of another additional language than L1 procedural knowledge (Falk, Lindqvist, & Bardel, 2013).

In addition, learners combine L1 and L2 knowledge and develop enhanced awareness or metalinguistic skills (Cenoz, 2003; Jessner, 2012, 2014). For example, bilingual learners outperform monolingual learners of English in a Basque context (Cenoz & Valencia, 1994). Evidence for increasingly beneficial effects of multilingualism suggests that the human mind can merge L1 and L2 knowledge while maintaining performance in each distinct language. For example, learning to pronounce new sounds can affect pronunciation of already established L1 sounds (Chang, 2012). Combining L1 and L2 will facilitate the mind to cope efficiently and flexibly with possibly redundant representations (Kovács & Mehler, 2009). This means that current findings predict facilitative effects of being multilingual in general. However, it is unclear to what extent there is a general language-independent multilingualism factor or whether this factor is composed of a summation over the specific characteristics of the languages learned. In addition to facilitative effects on additional language learning, being bi-/ multilingual facilitates efficiency of attention mechanisms and cognitive control (Costa et al., 2009; Costa, Hernández, & Sebastián-Gallés, 2008). However, it is unclear whether such general cognitive benefits enhance additional language learning.

Linguistic Distance

We hypothesize that L1 and L2 linguistic distances affect L3 learnability. The degree to which the L2 (or its absence) facilitates or impedes learning of a specific additional language can be defined as L3 learnability, and can be estimated through proficiency scores of L3 learners. Linguistic distance measures the degree of similarity between languages, which is often used for language classification (Ruhlen, 1991; Trask, 2000). Distance measures that fixate on the qualitative difference at the level of family and genus are available as well as quantitative distance measures based on the degree of linguistic differences (Greenberg, 1956; Nichols, 1992). Qualitative, language-family based notions of linguistic distance (Lewis, Simons, & Fennig, 2013) are useful for language learning studies in which a small number of languages are compared. For example, a study of the influence of Basque (as L1 or L2) vs. Spanish (as L1 or L2) on English as L3 (Cenoz, 2001) shows that a Basque background has a less positive effect on learning English than Spanish, irrespective of its status as L1 or L2. The exact quantitative linguistic distance between Basque and English is not straightforward to measure, but it seems obvious that Basque, an isolate language, is more distant from English than Spanish is, as English and Spanish are Indo-European languages.

For a comparison across a large number of L1s and L2s, a quantitative measure of linguistic distance is required to determine the effects of linguistic differences. Semi-quantitative measures can be used, which are based on the number of levels of the family tree the languages share (Adsera & Pytlikova, 2012; Desmet et al., 2009; Isphording & Otten, 2014), but such crude measures cannot distinguish between the similarity of Spanish and French to English, or Basque and Chinese to English. We therefore focused on linguistic data to provide measures of linguistic distance with more detail than is possible by counting nodes in language family trees.

Recently, typological resources have become available that are useful for measuring linguistic distances. Currently, researchers are starting to discover how linguistic data can be used to measure distances accurately. For example, basic vocabulary word lists (Dyen et al., 1992) have been used to statistically estimate the most likely time depth of Indo-European languages based on models of evolutionary language change over time (Bouckaert et al., 2012; Gray & Atkinson, 2003).

Similarly, basic vocabulary data for multiple language families (Holman et al., 2011) have been used to reconstruct language family trees based on the number of shared lexical forms. Both approaches compute distance measures between languages based on lexical differences. In addition, structural data (Dryer & Haspelmath, 2011) have been used for reconstructing family relationships (Dunn et al., 2005) and linking development of morphological differences across language families to changes in social structures (Lupyan & Dale, 2010). As adult learners experience problems with additional morphology (Ionin & Wexler, 2002), languages that are learned relatively often by adults may show gradual reductions in morphological complexity over time (Trudgill, 2011).

Here, we use both lexical and morphological distance measures to predict L3 learnability. For lexical distance measures, we use measures of evolutionary change within the Indo-European language family (Gray & Atkinson, 2003). We also use morphological distances, as these overcome the limitation of lexical distance to one language family only. Morphological distance measures are available for languages from non-Indo-European language families as well. Schepens et al. (2013a) used 29 morphological features extracted from the World Atlas of Language Structures (Dryer & Haspelmath, 2011) to construct three measures of morphological distance. They measured morphological similarity between Dutch and other languages, the degree of increasing morphological complexity from the perspective of a particular language towards Dutch, and the degree of decreasing morphological complexity from the perspective of a particular language towards Dutch. Both increasing morphological complexity and morphological similarity were significantly better predictors for explaining variation in speaking proficiency scores than decreasing morphological complexity. Increasing morphological complexity could replace morphological similarity without losing explanatory value, which explains why we decided to use the complexity measure. The complexity measure only takes into account the linguistic differences for which Dutch is more morphologically complex than the L1 or L2.

In contrast to the lexical distance measure, the complexity measure is asymmetric. Therefore, the increase in morphological complexity from Chinese to Dutch is not necessarily the same as the increase from Dutch to Chinese.

Chapter 3 (Schepens et al., 2013a) already established both L1 morphological distance effects and L1 lexical distance effects in a study which included both monolingual and multilingual learners. We expect that lexical and morphological measures together cover a full range of distance effects from very distant to very similar L1s and L2s in multilingual learners of L3 Dutch, including Indo-European as well as non-Indo-European languages. For example, lexically similar languages can differ in morphological complexity (e.g. English – German). Besides lexical and morphological differences, languages differ in syntactical (Dunn et al., 2011) and phonological ways (Moran & Blasi, 2014). The study of syntactical and phonological differences is, however, beyond the scope of the present investigation.

Large Scale Studies of Speaking Proficiency

Large scale studies of speaking proficiency have to deal with four challenging issues: 1) the difficulty of accurately and validly measuring language proficiency, 2) the availability and richness of information of learners' backgrounds, 3) the number of observations, and 4) the cross-classified nature of languages and countries.

This study makes use of results on the official state exam of *Dutch as a second language* (STEX) to study speaking proficiency. In the past, large-scale studies used self-reported proficiency measures (Hakuta et al., 2003). One of the reasons why these can be biased is that learners compare themselves to each other (Finnie & Meng, 2005; McArthur & Siegel, 1983; Siegel, Martin, & Bruno, 2001) and can thus systematically over- or underestimate their second language skills. In contrast, STEX is a database of proficiency *testing* scores. STEX provides testing scores that have been collected through a formalized judging protocol, while at the same time maintaining the large-scale availability of measures that is characteristic of census data. In 1989, a study reported correlations of .52 between objective measures and self-reported responses (Kominski, 1989). It was already noted then that a more objective measure of proficiency would be desirable to assess discontinuities in age effects on proficiency. Correlations between self-reported measures of proficiency and quick objective measures of proficiency are not strong and vary across groups, e.g. .5 for Dutch learners of English and .3 for Korean learners of English (Lemhöfer & Broersma, 2012). STEX aims to objectively measure the communicative competences of learners of *Dutch as a second language* at the B2 level of speaking proficiency of the Common European Framework of Reference for Languages (CEFR) (Hulstijn, Schoonen, de Jong, Steinel, & Florijn, 2012).

The STEX database is large enough to cope with the second and third issue, as speaking scores are available for more than 50,000 learners with information on language background and key individual characteristics. This is a relatively high number as compared to similar studies (e.g. the German Socio-Economic Panel analyzed by Isphording & Otten, 2011). Issue four, however, is quite challenging because we have to deal in our case with the cross-classified nature of not just the L1s and countries of birth but also the L1s and L2s. The L1s and L2s occur in many, but definitely not all possible language combinations. Addressing this issue is important for dealing with the fractionalization (degree of cross-classification) of languages and countries (Fearon, 2003). For example,

Methods

STEX

The STEX data come from the state exam Dutch as a Second Language (Nederlands als Tweede Taal, NT2). Passing this exam is a formal entry requirement for Dutch universities and for starting many higher level education jobs. The Dutch Governmental Board of Examinations provided exam results collected over a period of 15 years. The full state exam consists of a speaking, writing, listening, and reading test. The exam aims at a B2 passing level, which is upper-intermediate according to the Common European Framework of Reference (Council of Europe, 2001). The exam is mostly taken by a heterogeneous group of newcomers (150 different countries of birth). A second, alternative, state exam aims at a B1 passing level, and is meant for non-academic contexts. The scores from this exam are not analyzed here. Note that the exam requires considerable personal investment of time and money, which ensures high motivation of the candidates.

Sample

The study includes the first administered speaking proficiency scores for 39,300 multilingual candidates who reported both a mother tongue and a best additional language besides their mother tongue on the questionnaire (mean age 30.2, median 29, 26,225 females, 13,075 males). All candidates participated in the speaking exam between 1995 and 2010 at various locations in the Netherlands. Candidates with a country of birth, L1, or L2 with less than 15 candidates available were excluded. We also excluded candidates who gave missing or invalid (e.g. unreadable) answers to the questionnaire that candidates could fill in voluntarily before the exam. This resulted in the exclusion of participants who did not report age of arrival, country of birth, mother tongue. It was possible to determine linguistic distance for most of the L1s of the candidates included in our study. Out of 50,500 candidates, 11,200 candidates, including the monolinguals, had to be excluded because L1 and / or L2 linguistic distance could not be determined for them (in particular morphological distance, see Variables section). However, we compare scores for monolinguals with the scores for multilinguals to test for general effects of speaking an L2.

The multilingual candidates speak 56 different L1s (mean number of speakers 701.8, median 256.5) and 35 different L2s (mean number of speakers 1122.8, median 64, English representing 68.0%). Following WALS (Dryer and Haspelmath, 2011), the 56 L1s come from 32 different genera which belong to 14 language families. Of these languages, 27 are non-Indo-European and 29 are Indo-European (see

Appendix B for an overview). The candidates were born in 119 different countries from all over the world. Candidates have one out of 536 L1-L2 combinations (161 combinations had at least 15 candidates). 25.2% of all L1 speakers have an L2 other than the most common L2 for that L1 (which is mostly English), illustrating the cross-classified nature of the data. When candidates with English as an L2 are excluded, this value would be 38.0%.

Task

Candidates performed a mix of short and long speech tasks in 30 minutes (a typical exam consists of 14 tasks), in which they needed to provide or ask for information, give instructions, etc. For example, in the 1997 exam candidates needed to describe and give a motivated opinion about marketing campaigns in two minutes. The test assesses whether candidates can respond adequately to a given situation. The instructions were simultaneously provided through headphones and on paper. The performance tasks require the candidates to produce functional language. Performance is assessed in an experimental setting by placing candidates in a soundproof booth.

Measures

Standardized performance measures were computed using an item response theory model that enables comparison of test scores over time. The composition of the exams aims at measuring proficiency at the highest reliability level around the passing level of 500 points. The exams are less sensitive to differences around the maximum of 900 or the minimum of 100. Two independent examiners evaluate the spoken language on both content and correctness. The most important content criteria are the fit of the content to the task (about 30%) and the size of the vocabulary (about 18%). The most important formal linguistic criteria are word and sentence formation (word order, verbal inflection, tense; about 28%) and pronunciation (about 12%). These percentages are based on a speaking exam from 1998 but are representative for a typical speaking exam. The remaining 12% of the criteria are related to fluency, coherence, word choice, tempo, and register. In all, almost all criteria are influenced by lexical and morphological characteristics. Both professional examiners gave between two and six ratings for each task, depending on the duration of the task, about 40% being two-way (insufficient or sufficient) ratings and 60% being four-way ratings (insufficient, almost sufficient, sufficient, and good).

This study uses administered data for participants' country of birth, date of exam, date of birth, and gender, and questionnaire data for years of education, date of arrival in the Netherlands (useful to infer length of residence and age at arrival), L1, and L2. The question for the L1 was "what is your mother tongue", and the question for the L2 was "Do you speak another language besides Dutch and your mother tongue?" and, if the person answered yes, "Which other language do you speak? If you speak more than one, name the language that you know best" (literally translated from Dutch by this study's first author). Note that "best" is not quantified in our study and probably covers varying proficiency levels in the reported L2s. The correct interpretation of the Dutch form of the expression "Do you speak another language" is whether one knows how to express oneself orally in another language. In general, it can be assumed that learners already know this interpretation at the A2 level, which is below the passing level of the current exam. For example, the A2 level describes that a learner is able to "understand sentences and frequently used expressions related to areas of most immediate relevance."[10] Clearly, the A2 level requires candidates to acquire at least part of the target morphology.

[10] http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf p.24

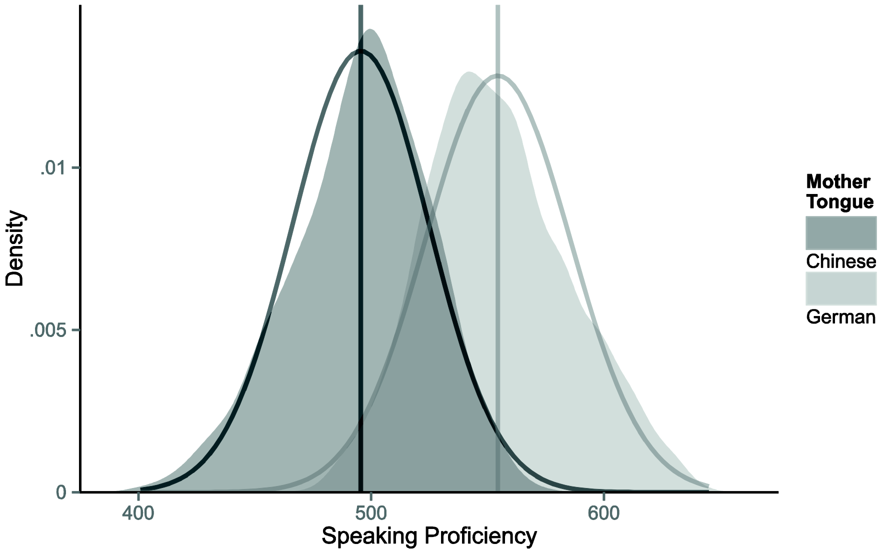


Figure 1. The distributions of German (N = 4773) and Chinese (N = 776) speaking proficiency scores compared with normal distributions, showing differences across subgroups.

Figure 1 illustrates the distributions of speaking proficiency scores using the Chinese and German native speakers, including monolinguals. German is the language most close to Dutch (besides Afrikaans); Chinese has a maximal lexical and high morphological distance. The graph shows that a native speaker of Chinese is unlikely to obtain higher scores than the average native speaker of German.

In Figure 2, three relatively prestigious Indo-European languages are crossed as L1 and L2 and Farsi was added as the fourth Indo-European language to visualize the effect of the three other languages as L2. The figure explores differences in L3 proficiency across the L1s and across the L2s without controlling for gender, age, education, and exposure effects. The differences between the L1s seem to be larger, but the L2 seems to matter as well, with L2 German producing the highest outcomes. The monolinguals have the lowest scores.

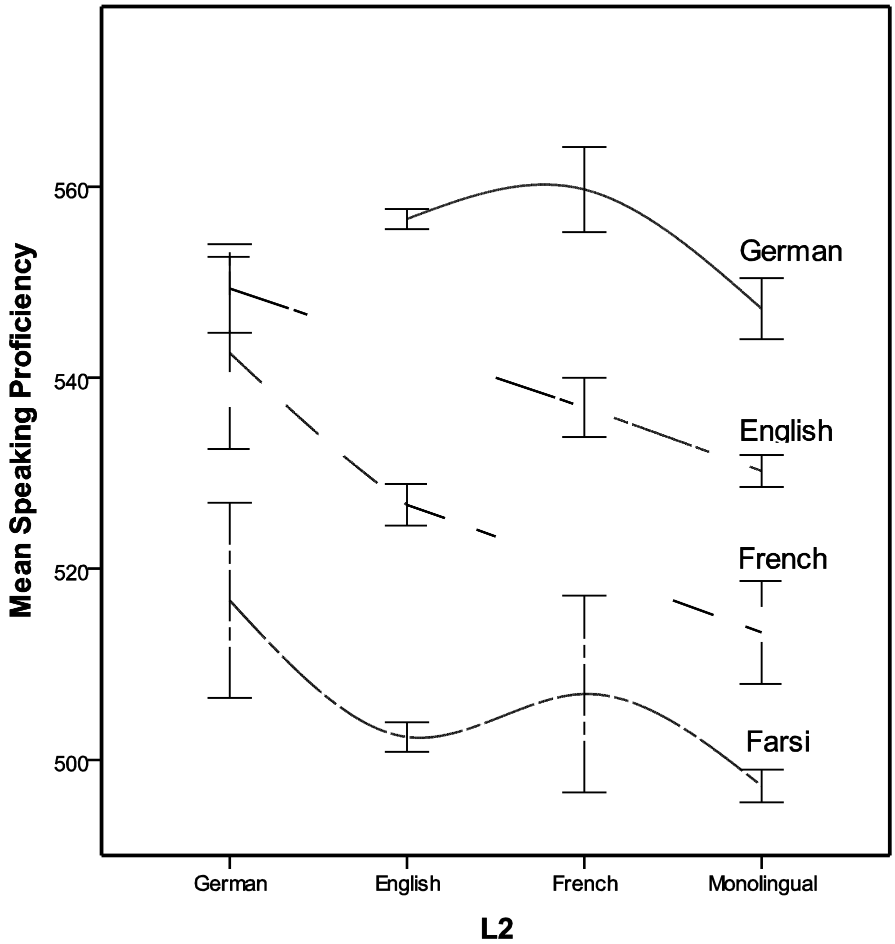


Figure 2. Mean speaking proficiency scores for 13 L1 – L2 combinations with 95% confidence intervals, all four L1s being Indo-European languages. The x-axis distinguishes between monolingual L1 and several multilingual L1-L2 combinations. The interpolations between mean speaking scores show downward trends, the monolinguals having the lowest scores.

We added measures of schooling quality, lexical distance, and morphological distance in order to explain variation across L1s, L2s, and countries of birth. Schooling quality was measured as the gross secondary school enrolment in 2006 (UNESCO, 2011), which is the ratio of total enrolment in secondary education per country (see also Schepens et al., 2013b).

Linguistic distance from the L1 and the L2 to Dutch was measured as the degree of evolutionary change based on shared cognates between Indo-European languages as measured in the sum of branch lengths that connect both languages to each other in a phylogenetic language family tree (Gray & Atkinson, 2003). This measure can be qualified as lexical distance. A maximum lexical distance as observed in the Indo-European language family tree was used for L1s from different language families. The lexical distances have higher values for larger distances: 0 means that the languages share all words in the Swadesh list and a higher distance means that the languages share a longer branch length, effectively having a lower lexical overlap. Lexical distances ranged from .0105 to .595 with a mean of .322 and a standard deviation of .175

Our second measure of linguistic distance was based on increasing morphological complexity. This measure of morphological distance was computed by comparing the complexity of Dutch to the L1 for 29 morphological features (Schepens et al., 2013a). For example, Dutch marks the feature “past tense” morphologically, whereas some languages don’t (Dahl & Velupillai, 2013). The distance measure is a weighted sum of these feature differences. Among these feature differences were differences in verbal person and number marking, the past tense, polar question coding, the question particle, coding of negation, inflectional synthesis of the verb, degree of inflectional morphology, etc. In contrast to the lexical distance measure, this measure also varies across non-Indo-European L1s also. Due to missing feature values in the WALS data, which was used to develop morphological distance, some missing morphological distance values were set to the same distance as linguistic neighbors (Bosnian to Croatian, Ukrainian, and Belarusian to Russian, Catalan to Spanish, Fulani to Wolof, Malay to Indonesian). The morphological distances have higher values for larger distances: 0 means of equal or higher complexity and the maximum distance (for Igbo) means of relatively lower complexity. Morphological distances ranged from -.017 to .327 with a mean of .050 and a standard deviation of .057.

We tested but left out after extensive model comparisons: geographical distance (from the capital of the country of birth to Amsterdam), the Greenberg diversity index (Greenberg, 1956)[11], same / different genus and / or family (Adsera & Pytlikova, 2012), peer group size (number of learners from country of birth taking part in the exam), total number of citizens in the country of birth, GDP per capita. These factors did not significantly influence L3 speaking proficiency testing scores.

Analysis

We used a mixed effects regression approach to predict variation in L3 proficiency scores across first and second languages. The approach summarizes over individual differences, resulting in aggregate scores across L1s and L2s that are controlled for third factors such as age, exposure, education, gender, differences across countries, and differences resulting from specific L1-L2 combinations. In the remainder of this section on statistical modeling, we describe the details of our approach. Note that a high degree of familiarity with mixed effects regression is necessary to understand the inferences that this approach allows.

The speaking scores were analyzed by using cross-classified random effect models (CCREM) in R (R Core Team, 2013) and fitted with *lme4* (D. Bates et al., 2014), which is a package that is not part of generic R. All candidates, irrespective of difference in language background, were included in one CRREM analysis by treating country of birth, L1, L2, and L1-L2 combinations as random effects (Schepens

[11] The degree of fractionalization between individuals across groups (languages) in an area (countries) affects political and economic developments (Desmet, Ortuño-Ortín, & Wacziarg, 2012). We tested for such effects by using Greenberg's measure of linguistic diversity (Greenberg, 1956) in the country of birth (number of languages per area).

et al., submitted). We keep this random effect structure constant throughout the rest of the chapter (except for the final analysis).

Effects of country background on L3 learnability need to be disentangled from language specific influence (Fearon, 2003). As many countries in the world are to some extent fractionalized in terms of languages, the most frequently used language in a country is not necessary the L1 or even the L2 of a candidate. Furthermore, inter-country linguistic differences can reach the level of completely different language families. A generalization from the country level to the language level is likely to neglect any existing linguistic diversity. We chose *lme4* instead of *nlme* (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2013) as *lme4* can fit models with partially crossed random effects in large unbalanced data, which is necessary given that not every language is spoken in every country.

CCREMs assume independence between crossed random effects (Baayen et al., 2008). The inclusion of random effects for both languages and countries only associates variation in proficiency scores with either languages or countries if that variation is unambiguously associated with a language or country, assuming that by-country and by-language variation is independent. This random effect structure results in conservative lower bound estimations of by-L1 variation and by-country variation (the minimal unexplained by-L1 and by-country variance that can be observed in the data).

It is often a mistake to assume that the random effects in cross-classified models are completely mutually independent in unbalanced data. The interdependency between languages and countries is not further investigated here. However, the degree of interdependency between L1s and L2s is potentially important for establishing whether L1 and L2 distance effects are indeed additive and independent. This is the reason why we also include L1-L2 combinations as an L1-by-L2 random interaction effect. In order to further control for dependency across the random effects, we decided to include random slopes for L1 linguistic distance across L2s and L2 linguistic distance across L1s. However, such models did not converge, probably because of the sparse crossing between L1s and L2s. Random slopes for the distance effects across countries also did not result in converged models.

We tested fixed predictors using a semi-backward elimination procedure (Baayen et al., 2008), in which we always retested a predictor after another predictor had been removed. We performed model comparisons with likelihood ratio tests as well as AIC comparisons for nested models. We computed evidence ratio's based on the AIC for non-nested models (Spiess, 2013). The higher the evidence ratio, the more evidence for a particular model there is (in favor of the model that is being compared to another model). We removed 787 outliers by excluding multilingual candidates who had a standardized residual on the speaking proficiency measure higher than 2.5 standard deviations from 0, amounting to 2.0% of the data. Residuals were based on a mixed effects model applied to the set of both mono- and multilinguals.

The intra-class correlation in a null model with no fixed effects (Goldstein, 2011) for the multilingual candidates (dividing the variance component of interest by the sum of variance components) indicates that 21.1% of the variation in proficiency scores is due to differences across L1s, 6.1% across L2s, 5.3% across L1-L2 combinations, and 18.0% across countries. 49.5% of the variation is due to individual differences. The null-level variance components were L1-L2: 3.07 (2.08, 4.17), Countries: 10.42 (8.85, 12.31), L1: 12.31 (9.76, 15.46), L2 3.51 (2.41, 5.14), residual: 28.75 (28.55, 28.96). We computed percentages of variance explained by subtracting the relevant variance component from the null-model variance component and dividing again by the null-model variance component (Kreft & Leeuw, 1998; Snijders & Bosker, 2011).

The variation across combinations may potentially be explained by interactions between L1 and L2 knowledge. Our goal is to reduce the language-related variances, while assuming that individual variation is homogeneous within each specifically observed language background. We already know that linguistic distance interacts with individual level factors such as age of arrival and length of residence from Chapter 2 (Schepens et al., 2013b). The effect of distance was higher for older learners and for learners who had resided in the Netherlands for a longer period of time. Although these cross-level interaction effects increase model fit, we observed that they reduce individual level variation rather than language-related variation.

Results

We start with testing L1 distance effects in all learners together (mono- and multilingual) and in multilingual learners only, in order to assess whether multilingual learners have higher proficiency scores than monolingual learners. We then proceed with the multilingual learners only, testing L2 language background effects on L3 performance. We compare the additive L1 + L2 distance model and two non-additive models. One non-additive model is based on the lowest distance of the L1 and L2 (typological primacy), the other one based on unique L1xL2 combinations (the L1xL2 interaction effect). Finally, we test whether L1 and L2 distance effects are robust against including both monolingual and multilingual learners in the same model and how this inclusion can be carried out.

L1 Distance Effects in Monolingual and Multilingual Learners

Chapters 2 and 3 (Schepens et al., 2013a, 2013b) showed, for both monolingual and multilingual learners, that lexical and morphological distance correlate with L2 / L3 speaking proficiency scores. Our first goal here, before we advance to an L2 distance model of by-L2 speaking variation, is to assess whether the L1 distance measures should still be included in a model of by-L1 speaking variation for multilinguals. L1 lexical and morphological distances are different dimensions of the communicative competences of L3 learners of Dutch. Therefore, both measures might explain part of the variance, which should become evident from likelihood ratio tests. According to Schepens et al. (2013a), an increased morphological complexity of an

L2 compared with an L1 results in lower L2 learnability; see also Linguistic Distance in the Background section above.

We first wanted to know how much of the remaining variation in speaking proficiency results from differences between monolinguals and multilinguals after accounting for L1 morphological and lexical distance effects. We started by testing whether there is an overall significant difference in proficiency between the monolingual learners of Dutch as an L2 (coded as 0) and the multilingual learners of Dutch as an L3 (coded as 1). This binominal variable is highly significant (B = 8.97, SE = .417, df = 47910, t = 21.526, p < .0001; $\chi^2(1) = 461.32$, p < .0001). The dummy variable shows that the L2 matters beyond the L1, and has an additional positive effect on the speaking proficiency score (almost 9 points) in favor of multilinguals.

Before testing L2 lexical and morphological distances, we tested whether the L1 distance effects remain the same when we restrict our analysis to multilinguals only. As was the case for the whole group, we found that L1 morphological similarity can be removed from a model with morphological similarity, increasing complexity, and lexical distance without reducing model fit significantly ($\chi^2(1) = .31$, $p = .5787$). Removing decreasing morphological complexity did not change the model fit significantly either ($\chi^2(1) = .02$, $p = .8912$). What remains is a joint model in which deleting both increasing morphological complexity ($\chi^2(1) = 13.21$, $p < .001$) and lexical distance ($\chi^2(1) = 21.99$, $p < .001$) decrease model fit significantly. This "L1 model" explains 47.7% of the by-L1 variance in speaking proficiency scores, which is more than either lexical distance (39.6%) or morphological complexity (30.8%) alone. Explained variance across countries of birth (43.7%), L2s (-4.1%), L1-L2s (2.4), and the individual level residual variation (2.3%) remained constant when either lexical distance or morphological complexity were removed. Note that a negative value for the proportions of such predictor-specific R^2 values can arise as a side effect from subsequent updates to a model in which variance is relocated to other predictors (Nakagawa & Schielzeth, 2013; Raudenbush & Bryk, 2002; Snijders & Bosker, 2011). The L1 model included gender, age of arrival, length of residence, the interaction between years of education and educational quality, L1 lexical distance, and L1 morphological complexity. All fixed effects are significant at the .001 level (except years of education, although its interaction with educational quality is significant) using Satterthwaite approximations, which are used to determine the effective degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2014). See the first four columns in Table 1 for estimates and confidence intervals for the factors in the L1 model for multilinguals. The last three columns show the "L1+L2 model" for multilinguals, which will be discussed in the next section. Interestingly, the directions of the effects and the effect sizes in Table 1 did not change as compared to analyses that include both monolinguals and multilinguals. Indeed, the L1 distance effect is present in the subset of multilinguals as well as in the group that included both mono- and multilingual learners.

We conclude that increasing morphological complexity can be used as a morphological distance measure jointly with L1 lexical distance and that they mutually complement each other in explaining variation across L1s in multilinguals. The more morphologically complex and the more lexically distant Dutch is compared to the L1 of the learner, the lower the proficiency.

Table 1. Estimates and confidence intervals for the random and fixed effects included in the L1 model and L1 + L2 model fitted to the multilingual learner group. The CIs are based on the profile likelihood.

	L1 Model			L1+L2 Model		
Effect	Estimate	2.50%	97.50%	Estimate	2.50%	97.50%
Random L1-L2 Variance	3.01	2.11	4.02	3.07	2.14	4.07
Random C Variance	5.87	4.80	7.06	5.89	4.81	7.07
Random L1 Variance	6.45	4.61	8.21	6.35	4.54	8.12
Random L2 Variance	3.65	2.54	5.31	2.34	1.34	3.63
Residual Variance	28.10	27.93	28.33	28.10	27.93	28.33
Intercept	527.45	518.64	536.09	534.26	525.1	543.46
Gender (Female = 1)	7.31	6.63	8.00	7.31	6.64	8.00
Age of Arrival	-0.65	-0.7	-0.61	-0.66	-0.7.0	-0.61
Length of Residence	0.53	0.46	0.61	0.54	0.46	0.61
Education Years	-0.09	-1.12	1.00	-0.10	-1.16	0.96
Education Quality	0.15	0.09	0.22	0.15	0.09	0.21
Education Years x Quality	0.03	0.02	0.04	0.03	0.02	0.04

L1 Morph. Distance	-63.19	-96.23	-30.69	-62.20	-94.87	-29.73
L1 Lexical Distance	-39.51	-54.25	-24.54	-40.22	-54.94	-25.5
L2 Morph. Distance				-18.52	-51.46	14.23
L2 Lexical Distance				-14.14	-23.56	-4.49

Note. Both distance measures are on continuous scales, lexical distances range from .0105 to .595 with a mean of .322 and a standard deviation of .175, while morphological distances range from -.017 to .327 with a mean of .050 and a standard deviation of .057.

Adding L2 Effects for Multilingual Speakers

The lexical and morphological distance measures are also suitable for measuring the lexical and morphological distance between L2s and Dutch. The question is whether lexical and morphological distance can explain variation across L2s as well as across L1s.

Likelihood ratio tests indicate that the updated model including both L2 lexical and morphological distance provides significant improvement of fit to the data ($\chi^2(2) = 15.32$, $p < .001$). Individually, lexical distance ($\chi^2(1) = 14.16$, $p < .001$) as well as morphological distance are significant as well ($\chi^2(1) = 7.30$, $p < .001$), although the improvement in fit for morphological distance is considerably smaller. The L1 + L2 model raises the explained by-L2 variance from -4.1% to 32.1%. Interestingly, lexical distance is largely responsible for the explained variance across L2s (32.4%), while morphological distance explains only 12.3% of the variation. Similarly, L2 morphological distance turns out to be non-significant in terms of the confidence interval estimates, see Table 1, although improvement in model fit is significant. Figure 3 shows the partial effects, contingent on the other predictors in the model. The model slightly increases the explained by-L1 variance in speaking proficiency scores (to 48.3%), while the individual variance and by-country variance remains the same. The explained variance across L1-L2 combinations is reduced from 2.4% to 0.7%, indicating that L2 distance does not increase the explained variance across L1-L2 combinations (it even decreases slightly).[12] Before the L2 distance effects are added to the model, the by-L2 adjustments for all 35 L2s correlate significantly with both lexical ($r = .58$, $p < .001$) and morphological distance ($r = .42$, $p < .05$). The correlations vanish completely when distance is accounted for in the statistical model. The remaining L2-adjustments are most positive for speakers of L2 German, Hindi, and Armenian, who may experience additional benefit beyond linguistic distance, and lowest for speakers of L2 Italian, Russian, and English, who may experience impeding effects despite favorable linguistic distance, see also Figure 4.

We conclude that L2 lexical and morphological distance both play a role but that the role of L2 morphological complexity is less strong than that of lexical distance given the current distance measures. Most importantly, the L1+L2 model predicts that there is an independent, constant L2 effect irrespective of the L1. It means that the added value of L2 French, for instance, is constant, irrespective of whether the L1 is German or Spanish.

---

[12] A number of different general R^2 estimates (in contrast to variable-specific measures) of goodness-of-fit are available for mixed effects models (Nakagawa & Schielzeth, 2013), we report two of them here. The overall variance explained by all of the fixed effects is 58.1% (general R squared marginalized over the random effects). The variance explained by the fixed effects, given that the random effects are known, is 91.7% (simple R squared, conditional on the random effects). The AIC as estimated from an unrestricted maximum log likelihood fitted L1 + L2 model is 366,963.0 versus 366,974.4 for L1 only. This is a clear improvement in model fit (Baayen, 2008).

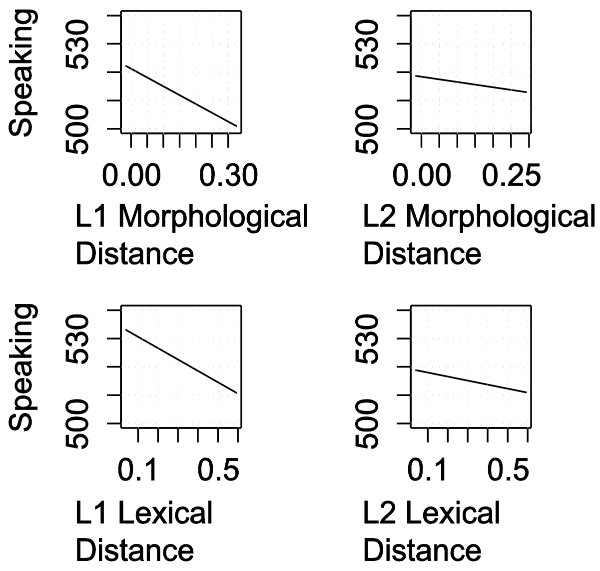


Figure 3. Partial effects for the L1 + L2 model. The L1 has a stronger influence on learning L3 Dutch than the L2 has.

A Non-Additive L2 Effect: Typological Primacy.

A currently outstanding problem in the literature is whether the status of having learned a language as L1 or L2 determines whether L1 or L2 knowledge will be transferred to the L3. It has been proposed that knowledge from the typologically closer language is more likely to be transferred than knowledge from the more distant language, irrespective of having learned the language as L1 or L2 (the status factor). If an L3 learner speaks an L2 that is typologically close to Dutch, the L2 may replace the primary L1 effect and the L1 effect may or may not still behave as a secondary effect. We will refer to this new distinction as the “min model”. We can fit the min model by selecting the minimal distance (min) of both the L1 and L2 distances. The minimal distance is the distance to Dutch from the language least distant to Dutch. Its opposite, the maximal distance, is the distance to Dutch from the language most distant to Dutch.

The evidence for the min model is more than one million times lower than the evidence for the L1 + L2 model (using an evidence ratio test based on AIC; Spiess, 2013). This indicates that the L1 and L2 effects are more stable across learners than minimal distance effects. The min model can be improved by adding the maximum distance to it ($\chi^2(2) = 31.62$, $p < .001$), suggesting a secondary effect for the most distant language of the L1 and the L2. However, the evidence for the max + min model is still 100,277.2 times lower than the evidence for the L1 + L2 model.

We conclude that the data do not support a min model or min + max model as strongly as they support an L1 + L2 model. The status of the L1 and the L2 is an important determiner of their influence, the status or impact of the L1 being more important than the status of the L2.

The L1 x L2 Interaction Effect

Although the additive L2 distances provide a significant improvement to an L1 distance only model, it is possible that the learners' L1 distance negatively interacts with their L2 distance. The unexplained variance across L1-L2 combinations in the null model was estimated at 5.3% of the total unexplained variance across learners. The L1 + L2 model only accounts for 0.7% of this percentage. Can a multiplication between the L1 distance and L2 distance explain a significant part of the random variance across L1-L2 combinations? We may expect that a highly similar L2 will provide a relatively higher benefit to speakers of a particular L1 if that L1 is relatively distant from Dutch. The same L2 may provide less of a benefit if an L1 is close to Dutch already. A positive interaction effect (L1 multiplied by L2) may be able to capture this. The closer the L1 and the L2 are to Dutch the more they will support each other; the more distant the L1 and L2 are to Dutch, the more they will diminish each other's added value. Cases in which either the L1 or the L2 is similar (and the other distant) will be relatively more beneficial than in the additive cases. For example, suppose that English and German are both 5 times more similar to

Dutch than French and Spanish are (e.g. distance of .1 for English and German versus .5 for French and Spanish). Then, in the interactive case, L1 Spanish – L2 English is 5 times less similar to Dutch than L1 German – L2 English (.1*.5 versus .1*.1), and 5 times more similar than L1 Spanish – L2 French (.1*.5 versus .5*.5). In the additive case, L1 Spanish – L2 English is 3 times less similar than L1 German – L2 English (.5+.1 versus .1+.1), but only 1.66 times more similar than L1 Spanish – L2 French (.1+.5 versus .5+.5).

To test for an L1 x L2 interaction effect, we added multiplicative terms between lexical and morphological L1 and L2 distance to the previous model. Likelihood ratio tests indicate that the "L1 x L2 model" does not fit the data better than the L1 + L2 model ($\chi^2(2) = .87$, $p = .6458$). Neither lexical distance ($\chi^2(1) = 0.40$, $p = 0.5264$) nor morphological complexity ($\chi^2(1) = 0.34$, $p = 0.5598$) were significant. The L1 x L2 model did not explain additional variance across L1-L2 combinations and the L1 x L2 multiplicative effect is neither in the 95% confidence interval, nor significant according to Satterthwaite approximations. We conclude that L1-by-L2 random interaction cannot be explained by an L1-by-L2 fixed multiplicative effect of either lexical or morphological distance.

It may be the case that uncommon L1 – L2 combinations produce unstable effects and that they obstruct our chance of observing systematic multiplicative effects. To assess whether the interaction of L1 – L2 combinations is also not significant in the most common L1 – L2 combinations, we removed L1 – L2 combinations that appear less than 15 times in the database. The resulting models show that multiplicative L1*L2 distance effects remain non-significant, irrespective of removal of uncommon L1 – L2 combinations, see Table 2.

Multiplications between different distances cannot account for the patterns observed across combinations beyond individual additive distances. The remaining variation across L1-L2 combinations is more complex than a simple multiplicative effect.

Effects of L2 Distance or Multilingualism?

It now seems evident that an independent L2 distance effect operates alongside an L1 distance effect. Part of this explanation, however, depends on the assumption that the addition of an L2 to the language inventory of the speaker does not affect L1 behavior, as the two effects are independent. If that is true, monolinguals speakers would have the distance advantage of their L1, but no profit of the L2 as there is no L2. We re-examined the L1 effect in the complete group of monolingual and multilingual speakers. Monolinguals were given the value “monolingual” for their L2. As L2 distance is not defined for monolingual speakers, we set L2 distance to the highest observed distance across all L2s (for lexical distance: Albanian; for morphological distance: Igbo). Setting the L2 distance to the L1 value would wrongly model that monolinguals benefit twice from their L1. We also added the dummy variable for being multilingual in order to test whether the general effect of being multilingual is still significant after L2 distance is accounted for.

We find that a dummy variable for multilingualism is non-significant when the L2 distance measure already accounts for by-L2 variation. However, the general effect of the multilingualism dummy variable is significant when random effects for the L2 and L1-L2 combinations are removed. This suggests that there is a gradual difference between monolinguals and multilinguals, depending on the type of L2. In a model including a random effect for the L2 instead of the dummy variable (and no L2 distance measures), the score adjustment for the monolinguals is the most negative one observed (see Figure 4), reflecting a (gradual) difference between monolinguals and multilinguals.

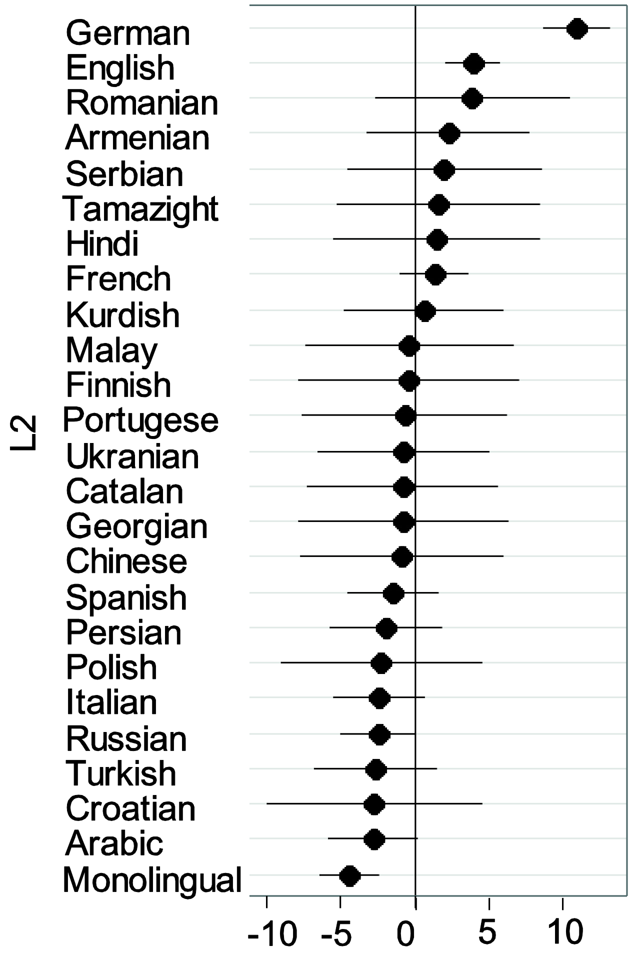


Figure 4. L2 score adjustments per L2 after L1 distances but before L2 distances are added to the model for the 25 most common L2s. The remaining random variance across L2s shows how monolinguals are hindered more than speakers of any L2 are, although the size of this difference depends on the L2.

Table 2. Comparisons of additive L1 and L2 distance effect models to models including L1*L2 multiplicative effects. Multiplicative effects for both lexical as well as morphological distance are included. The lower the evidence ratio, the less evidence there is for interactive effects.

L1-L2 < 15	Monolinguals	χ^2	df	p	Evidence Ratio
Included	Excluded	0.87	2	.65	<.001
Included	Included	1.60	2	.45	<.01
Excluded	Excluded	2.70	2	.26	< .0001
Excluded	Included	1.81	2	.40	< .001
Excluded	Included + Dummy	1.85	2	.40	< .001

We also find that including monolinguals does not change the estimations of the L1 and L2 distance effects. Explained variances remain robust against the inclusion of monolinguals in the L1+L2 model. Also, after adding monolinguals and setting their L2 to a maximum distance from Dutch, L1*L2 distances remain non-significant (see Table 2). We also compared the effects of including and excluding monolinguals and uncommon L1– L2 combinations on the distribution of variance across the random effects. This showed only small changes: the inclusion of monolinguals gives an increase of variation across L1 – L2 combinations while excluding uncommon L1 – L2 combinations results in a slight increase of explained variance across L1s, but not across L2s. We conclude that knowledge of an L2 can be helpful in general but also that its effect size mainly depends on the specific L2 to L3 distance.

Discussion

We investigated effects of L1 and L2 distance on L3 Dutch learnability in state examination data (STEX). We started by hypothesizing that the lower the L2 distance is, the higher the L3 learnability, and that this effect is weaker than the effect of L1 distance. Accordingly, robust additive L1 and L2 distance effects were found in upper intermediate learners of L3 Dutch, where the L1 effect was stronger than the L2 effect. Cross-classified random effect models were used to decompose variance in speaking proficiency test scores into by-country, by-L1, by-L2, and L1-by-L2 variation. L1 and L2 distances were measured with lexical and morphological distance measures. There were five main findings in our study:

1) Both lexical distance (39.6% across L1s and 32.4% across L2s) and morphological distance (30.8% across L1s and 12.3% across L2s) explain a significant proportion of the variance in the group including multilinguals only. Adding monolinguals to this model does not change these effects significantly.
2) There is more variation in L3 proficiency across L1s than across L2s (21.1% versus 6.1%). The L1 distance effect is stronger than the L2 distance effect (distances explain 47.7% of the variance across L1s versus 32.4% across L2s).
3) An additive L1 + L2 distance model fits the data better than more complex L1 x L2 multiplicative distances.
4) Language status (whether a language is an L1 or L2) fits the data better than an alternative model that gives primacy to a language based on the smallest distance.
5) Being multilingual is generally better for learning a new language than being monolingual, provided the L1 is the same. This facilitative effect also applies when the L2 has a large distance to the new L3 language. The effect is smaller for larger distances and not always more beneficial than being monolingual.

The observed additive L1 and L2 distance effects provide evidence for a theory of L3 learnability in which knowledge of previously acquired languages has to be accounted for. We discuss how these findings relate to existing theories of L3 learning, validity issues in measurements of L3 performance, and issues in L1 by L2 mixing.

Consistency with Existing Theories of L3 Learnability

We found that the L1 is more important than the L2. This differs from a model that predicts no special importance of the L1 (Flynn et al., 2004) and a model that predicts an L2 blocking effect to the L1 (Bohnacker, 2006). Both models are presented in Table 3. Our results nicely meet the predictions of the cumulative enhancement model (Flynn et al., 2004): (1) the effect of the L2 (1) is not absorbed by the L1, (2) is either neutral or positive, and (3) is more beneficial for learning an L3 than having no L2 at all. With respect to (3), we cannot reliably conclude that all the 35 L2s, including the distant ones, provided added benefit above and beyond having no L2 at al. Our outcomes add two additional conclusions: (1) L1 influence is stronger than L2 influence, and (2) the degree of L1 and L2 influence is related to the respective L1 and L2 distances. The degree of L1 and L2 influence is not affected by the distance of the other language involved and the L2 distance effect does not prohibit, impede, or enhance the L1 distance effect. L2 distance can benefit the learner, and this effect adds to and does not substitute or change the L1 distance effect.

What are the consequences for the other two theories mentioned in Table 3? Our findings align with a variation of the L2 status factor model (Bardel & Falk, 2007). We find that the status of a language has to be taken into account when accounting for the typological similarity in L3 learning. However, typological factors do not change the importance of the L1 (Rothman, 2011), even if the L2 is closer to the L3 than the L1.

Table 3. Predictions of studies discussed in the Introduction and findings from the present study.

Reference	Model	Prediction	Finding
Flynn et al., 2004	Increasingly cumulative	Bilingual > monolingual	L2 distance > monolingual
Bohnacker, 2006	L2 blocks L1 transfer	L2 > L1	L1 > L2
Bardel & Falk, 2007	L2 status factor	L2 status > L1 distance	L1 status > min distance (L1, L2)
Rothman, 2011	Typological primacy	min distance (L1, L2)	L1 + L2

Evaluation of L3 Performance Measures

We expect that additive L1 and L2 distances not only predict performance in the state exam Dutch as a Second Language, but that this finding bears on L3 learnability in general as well. With respect to generalization to L3 learnability, we discuss the specific combinations of L1s and L2s and the focus on Dutch as the L3, the measurement of proficiency scores, and the information on the level of L2 proficiency.

Throughout this chapter, we used the term L3 learnability to refer to the relative L1 and L2 influence on the proficiency of the candidates in L3 Dutch. Candidates may have any L1 and knowledge of one or more additional languages, representing a wide sample of languages across the world. Unlike English, French, Spanish, or German, Dutch is a language not often taught in secondary schools across the globe. This avoids a comparability problem because of L2/L3 exposure in country of birth. Although such problems may admittedly complicate the study of world languages like English or Spanish as L3s, we believe that the present study provides directions to conduct such research.

The passing level of the test we used is relatively high (B2 level) and requires that learners speak the L3 relatively easily. The wide variety in speaking proficiency scores and their normal distribution (see Figure 1) enables the detection of distance effects. The passing criteria are related to language proficiency only. General intelligence is certainly required to provide argumentation in some of the tasks. The assessment focuses on adequateness of argumentation in a given situation (levels of proficiency are defined in the Common European Framework of Languages). Lexical and morphological aspects are both significant parts of the judgment criteria. Form is almost as important as content. Sample exams (specimen exams) are published regularly and are accessible to the public. Most students make use of exam-specific training as such training is widely used in language learning classrooms. Local Dutch authorities are responsible for maintaining accessibility to the exam, which implies that language classes are subsidized. Besides language practice, language classes have an important societal participation component.

Linguists tend to be critical towards the accuracy and reliability of test scores (Munro, 2008). However, test scores are almost certainly more accurate and reliable than self-reported proficiency levels. If there is remaining noise in the STEX proficiency measures, the estimations of the distance effects reported here are at least lower bound conservative estimates of the actual distance effects. However, care should still be taken in interpreting the importance of the L2 compared to the L1, as the L2s reflected subjective judgments of candidates' best additional languages. We do not expect, however, that many participants report L2s in which they do not know how to express themselves at all. The self-assessment of knowing how to express oneself in an L2 is made in the formal setting of a B2 language exam, and self-reported proficiency is known to correlate weakly, but significantly, with more objective measures of proficiency. We cannot exclude the possibility that candidates reported a second language from which they cannot transfer anything due to low proficiency. Similarly, candidates have competing candidate languages. Theoretically, such noise, if present, only leads to an underestimation and not to an overestimation of L2 distance effects. More research is necessary to study whether a higher L2 proficiency leads to stronger L2 distance effects.

L1 x L2 Mixing

Additive L1 and L2 distance effects apparently provide a better understanding of how L1 and L2 effects work independently. The model has the benefit of being straightforward and transparent. The data does not provide support for multiplicative distance effects, but we need to be careful. Variance remains in our data in the mix of different L1 and L2 types that is not yet explained by additive L1 and L2 distance effects. It is unclear whether this variance in specific L1 x L2 pairs is due to intricacies in the data set under investigation, i.e. due to other variables not included in our database (e.g. certain social conditions), or due to particular supportive or impeding L1 x L2 pairs. For example, after looking specifically at L1-L2 combinations, we observed that the L1 + L2 model does not explain how L1 Polish interacts with L2 Italian, as observed performance is lower than the model's prediction. We found no evidence that the combination of the L1 and the L2 can explain such intricacies. This does not exclude the possibility that that social factors such as social class or the amount of working hours may be important.

Our findings suggest that L1 influence does not change by adding an L2, suggesting that speakers efficiently combine L1 and L2 influences on L3 learning. It means that the L1 effect in monolinguals remains comparable to the L1 effect in bilinguals, suggesting that L1 knowledge remains intact after an L2 has been added. Bilinguals make use of an additive L2 distance effect, which leads to performance increases in learnability.

Linguistic Distance

Linguistic distance nicely captures the regularity in L1 and L2 dependent learning difficulty of a third language, i.e. L3 learnability.

The variation in L3 learnability cannot be explained by only one distance effect; evidence for two additive effects of linguistic distance was found. We used lexical distance for distances to Indo-European languages with a maximum distance to non-Indo-European languages, and morphological distance for distances to both Indo-European and non-Indo-European languages. In addition, as learning difficulty is not necessarily symmetrical, we used lexical distance as a symmetric distance measure and morphological distance as an asymmetric distance measure. In combination, the measures account for asymmetric and symmetric distances within Indo-European languages, and asymmetric distances only within non-Indo-European languages. In addition, as L1 distance, age, and speaking proficiency interact, future work could examine the non-linearity involved in the role of L2 distances. Furthermore, since the phonology of an additional language has a persistent influence on L2 learning difficulty as well, we also intend to investigate effects of phonological distance on L2 and L3 learnability in the near future.

Conclusion

This study investigated whether linguistic distances of previously acquired languages predict the learnability of an additional language. The study shows that lexical distances within language families and morphological distances between languages from different language families can be used to answer these questions. The results demonstrate the importance and robustness of distance effects from additional language knowledge (L2) in learning a new language, the L3. That is precisely what was predicted by our main hypothesis. We can conclude that the closer the L2 is to the L3, the higher is the learnability of the L3. Learning an L3 becomes more difficult the more lexically distant and morphologically less complex the L1 or L2 is. Does that imply that an L2 can have a negative or impeding effect on learning a new language? Our second hypothesis was that the L2-L3 distance effect is facilitative, the lower the distance is, as multilingual learners obtain higher target language proficiency scores in Dutch than monolinguals. No significant effects of being multilingual remain after the L2 distance effects are accounted for, which indicates that general cognitive effects of multilingualism can be decomposed into independent L1 and L2 distance effects. The monolinguals only have L1 distance effects and do not profit nor are they hampered by additional L2 effects. Finally, the outcomes corroborate our third hypothesis that states that the L2 distance effect is weaker than the L1 distance effect.

Previous studies have not provided conclusive evidence for independent additive L2 distance effects, because of small-scale data or because of the lack of distance measures. A central step in our analyses was the definition of distances, both on the lexical and the morphological level. Evidence from large-scale comparative data across language backgrounds seems useful for the study of distance effects on L3 learnability. This approach can further benefit from studies of other languages besides Dutch. The frequency with which multilingualism occurs in society creates an opportunity for large-scale analyses of the persistent diversity in the learning of additional languages by adults. The present findings provide new empirical evidence for theories that predict distance effects of both the L1 and L2 that cannot be neglected in the learning of additional languages by adults. The present findings provide new empirical evidence for theories that predict distance effects of both the L1 and L2 that cannot be neglected in the learning of additional languages by adults. Our study does not predict concrete transfer phenomena. Instead, our study shows how the degree of effort necessary to learn a L2 or L3 varies depending on L1 and L2 linguistic distances. This implies that absence of specific forms of transfer is not counterevidence against global L1 and L2 linguistic distance effects. We hope that our findings will contribute to helping learners to bridge the distances that exist depending on their language background, by adapting and improving second and third language learning courses to the specific language combinations involved (Rivers & Golonka, 2009).

Chapter 7

Discussion and Conclusions

The foregoing chapters investigated the effect of linguistic distance on the learnability of Dutch as an additional language in adult learners with a variety of language backgrounds. The cover of this book pictures a page from the Voynich manuscript, which is a manuscript that seems to speak in tongues. It uses language-like signs with no apparent meaning. Listening for the first time to a new language may also arouse the bewilderment of speaking in tongues, but at closer examination, when learning the language, it becomes clear, slowly perhaps, that the sounds express meaning.

Language learners have to cope with differences between languages to bridge linguistic gaps. For example, native speakers of German who are learning Dutch have to discover that Dutch makes a distance distinction in its demonstratives, e.g. *deze muis* and *die muis* meaning either *diese Maus hier* or *diese Maus dort*. Despite the new sound /ui/ in *muis* (IPA /œy/), which does not exist in German, German learners can make safe use of the form similarity between *muis* and *Maus*. English learners of Dutch already make such a two-way distance distinction (*this mouse* and *that mouse*) in their demonstratives, giving them the opportunity to use both the word for mouse and the demonstrative distance distinction.

This thesis aims at developing measures of linguistic distance that can account for differences in learnability of Dutch as a second or additional language across a variety of language backgrounds. Language acquisition research on adults provides abundant evidence that the learnability of an additional language later in life is substantially lower and more variable than the learnability of an additional language early in life (Birdsong & Molis, 2001; Hakuta et al., 2003; Johnson & Newport, 1989). The considerable variation in L2 learnability among adult learners of Dutch, perhaps partly because of decreasing learning competencies, sheds light on how adult learners manage the relative challenges of language learning, as co-determined by their language background (Brown, 1998; Chiswick & Miller, 2005; Flynn, Foley, & Vinnitskaya, 2004; Ionin & Montrul, 2010; Isphording & Otten, 2013; Kellerman, 1995; Odlin, 2005; Ringbom, 2007). This thesis proposes concrete linguistic distance measures that are in useful in multilingualism research on cross-linguistic influence and transfer. We compared L2 Dutch proficiency scores for a large variety of language backgrounds and we developed measures of linguistic distance to explain divergence in L2 learnability.

In this chapter, we first summarize our results, then we discuss our approach, and we end by discussing the importance and implications of our findings.

Summary of Results

The Effect of Linguistic Distance

Chapter 2 showed that lexical distance affects L2 learnability. Lexical distance measures successfully explained around 75% of the variation in speaking proficiency scores across a large range of 35 Indo-European language backgrounds. We tested the effect for two lexical distance measures: a measure of language diversity in terms of the degree of evolutionary change and a measure of language disparity in terms of the percentage of shared cognates. We found that evolutionary distance better captures the relatively small differences from Dutch to Germanic L1s and the relatively large differences from Dutch to non-Germanic L1s, as compared to the shared cognates measure. The lexical distance effects were robust against interactions with age of onset, exposure, gender, and education. After controlling for these effects, we observed a correlation of .87 between lexical distance as based on evolutionary change and L2 learnability.

In Chapters 3 and 4, we proceeded with studies of more detailed morphological and phonological accounts of the variability in speaking proficiency scores across L1s. Chapter 3 established that distance has an effect in the morphological domain across 49 Indo-European and non-Indo-European languages. We found that the impact of morphological complexity on L2 learnability depends on the increase in complexity starting from the L1 to the L2 morphology. After controlling for third factors, we observed a correlation of .77 for morphological similarity, .67 for increasing morphological complexity, and .45 for decreasing morphological complexity. In general, complexity has many (subjective) dimensions (saliency, elegenace, etc.), even within the specific domain of morphology. Our sample of features includes 28 dimensions of morphological complexity. Some morphological features of the L1s were more complex than Dutch, while other were less complex. We combined the features into language-specific measures to characterize the general notion of morphological complexity in relation to Dutch. Likelihood ratio tests supported this notion of morphological complexity.

Chapter 4 addressed distance in the phonological domain, showing that phonological distance maps on variability in L2 learnability across 62 Indo-European and non-Indo-European languages. We compared phonological distance as based on new sounds and new distinctive features. We compared the 38 sounds of Dutch to the sounds in the sound inventories of the L1s of the learners. Besides counting the number of new sounds, we also compared the sounds in terms of their distinctive features. This featural representation is widely accepted to have psychological reality by many phonologists. The phonological "feature" has specific physiological properties. Our measure taps into this physiological way of representing phonological differences. L2 learnability was generally lower when new sounds had more distinctive features than their nearest neighboring sound in the L1. We found that the impact of the L1 sound inventory on L2 learnability depends on the number of new sounds in the L2 ($r = .35$) and the number of new features with respect to the nearest L1 sounds ($r = .47$).

Together, Chapters 2, 3, and 4 demonstrated that linguistic distance measures successfully explain variability in L2 learnability across different linguistic domains.

The Effects of L1 and L2 Distance

Chapters 5 and 6 focused on effects of linguistic distance in the additional language background of learners of Dutch as an L3. These chapters tested the effects of linguistic distance for language background that has an L1 status and a L2 status as well, where status indicates whether a previously learned language was learned as L1 or L2. Chapter 5 decomposed variation in speaking proficiency scores in 50,500 learners of Dutch. It showed that additional language background matters beyond the L1, but that the L1 accounts for a larger percentage of the variance. Additionally, Chapter 5 showed that the variation across L2s is robust against specific combinations of L1s and L2s. The model that fitted the data best attributes variability in L3 speaking proficiency scores to independent variance components for the L1 (20% of the variance), the L2 (5%), L1-L2 interactions (5%), country (20%), and individual differences (50%). The patterns of random by-L1 and by-L2 adjustments were similar ($r = .60$), suggesting that L2 variance also follows a distance ordering. Finally, this chapter made it clear that random interaction effects are useful to model dependency between random effects in cross-classified random effect models.

Chapter 6, building upon the variance structure as described in Chapter 5, showed that L1-L2 distance effects successfully explain variation across L1s and that L2-L3 distance effects successfully explain variation across L2s, irrespective of the inclusion of monolinguals. These findings align with the Cumulative Enhancement Model (Flynn et al., 2004). L1 distance explained more variation across L1s (50%) than L2 distance explained across L2s (30%). Moreover, we concluded that the L1 and L2 distance effects are additive: both the L1 and L2 matter independently. Chapter 6 also revealed that multilingual learners perform better than monolingual learners do, but that the beneficial effects of being multilingual could be explained by L2 linguistic distance.

General Summary of Results

In all, effects of linguistic distance on the learnability of Dutch as an additional language show up in different linguistic domains (Chapter 2, 3, 4), in different first and second language statuses (Chapters 5, 6), and in different learning stages and conditions in terms of age, exposure, gender, and educational effects (Chapters 2-6). The larger the linguistic gaps learners need to bridge, the lower the learnability. The more evolutionary change as based on lexical distance, the higher the "step up" in morphological complexity, and the more new features, the lower the L2 learnability. In addition, lexical and morphological distances also influence L3 learnability. Multilevel modeling enabled the comparison of the relative importance of L1 and L2 distance via decomposition of variance across L1s and L2s. Speaking proficiency scores allowed the comparison of different types of distance measures in different domains.

Approach

This section discusses the data and methodology. We start with a discussion of language testing and typological data, and we end with a discussion of statistical modeling.

Language Testing

We made use of a large-scale database of speaking proficiency scores from a unique longitudinal language-testing database of the state examination for Dutch as a Second Language (STEX). The Dutch Board for Exams (Board for Exams Act, see Staatsblad 2009, 93) currently runs the assessment of speaking proficiency according to the State Exam Dutch as a Second Language Act (see Staatsblad 1993, 569). The functions of this exam are discussed in Chapter 1. This section discusses the usefulness of language testing data for investigating distance effects in L2 learnability. A potential problem with language testing data is how variation in learner performance can shed light on the L2 learning process. Here, we discuss how expert examiners assess speech on intelligibility. Secondly, we discuss collinearity issues between overall speaking proficiency measures and lexical, morphological, and phonological linguistic distance measures, which make it difficult to tease different distance effects apart.

STEX scores of speaking proficiency are based on expert L2 proficiency ratings of a number of aspects of the spoken output according to formalized criteria. The testing procedure might result in a noisy measure of speaking proficiency, as it is not possible to assess all aspects of speaking proficiency in a test. In contrast to self-reported proficiency measures, however, it is unlikely that testing scores are similarly biased as the test is validated and maintained regularly in comparison. The validity of testing procedures, on the other hand, is questioned as well. It is not trivial, for instance, to tease accentedness and intelligibility apart (Munro & Derwing, 1995). Amongst others, a training of evaluators is supposed to help the correct judgment of foreign accentedness. Munro (2008) states that not much is known about how well examiners are capable of determining intelligibility for everyday language use. The examiners of the L2 Dutch exam are certified L2 Dutch teachers who have additionally passed a training and evaluating exam (Board of Examinations, 2011). Although effects of accentedness on examiners may occur, the criteria of the test and its portioning into specific ratings ensure specificity to intelligibility problems. The Dutch society of second language teachers regularly discusses L1-specific errors and accents. However, examiners' L2 background biases seem likely to affect the ratings they make (Winke, Gass, & Myford, 2013). On the other hand, other findings suggest that listeners rapidly adapt to a new foreign accent (Bradlow & Bent, 2008). Such adaptation processes may explain why examiners may be able to judge uniquely deviating foreign accents on intelligibility without having been exposed to such an accent before.

Thus, language-testing data provide accurate proficiency estimations for a large number of learners with many different language backgrounds over time. We do not argue that such data can replace carefully controlled linguistic studies of speaking proficiency, but we do think that evidence from such large-scale data sources provides an essential addition to experimental approaches. In particular, the approach differs from experimental approaches to L2 learning in terms of participant selection (e.g. (Henrich, Heine, & Norenzayan, 2010), small Ns, and high individual variation.

STEX contains L2 speaking proficiency scores for a wide range of different subpopulations. Figure 1 shows the differences in speaking proficiency scores on the y-axis and the variation across L1s, illustrating the use of a large set of data for calculating differences between L1s and their language families while controlling for third factors (age, exposure, gender, education) and implicitly for individual differences (e.g. tiredness, hours of study, etc.). Figure 1 is based on first-time speaking proficiency scores for 50,236 candidates as collected over a period of 15 years. Figure 2 examines the stability of the performances over four different L1 groups. The figure shows that the group differences fluctuate over time. As our data pertains to a period of 15 years, a time factor, e.g. in terms of norming differences, may be present. An explanation for a potential regression towards the passing level could be that the exam's questions concentrate more on the area surrounding the passing level. In other words, the precision of the test increases over time. Another potential interfering variable lies in the challenge to deal with the administrative complexity in a continuous manner. For example, one of the priorities of the administrative staff was to make sure that re-exams were administered as such and not as an exam of a new candidate. Analytically, if the scores from re-exams generally increase, effects of administrative errors of re-exams are assumed to lower estimations of linguistic distance effects instead of strengthening them. We did not find evidence for such errors in the data after examination of changes in the numbers of candidates per year. In all, the longitudinal and large-scale availability of testing scores allows aggregation over noise over time and individual differences in the data.

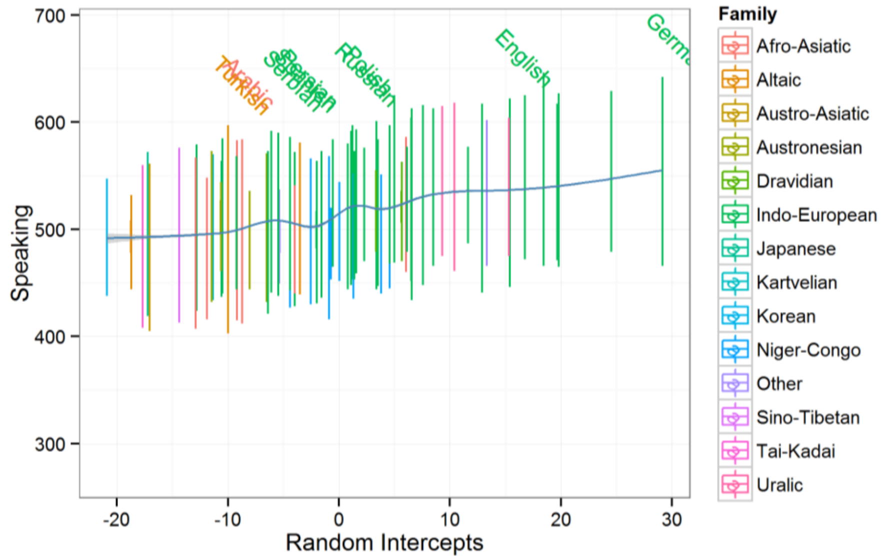


Figure 1. Variation across L1s and their language families, after controlling for age, exposure, gender, and education. The bars represent the range between all scores for a specific L1 that are not outliers. Outliers are values larger than the highest or lowest value that is within 1.5 * inter-quartile range.

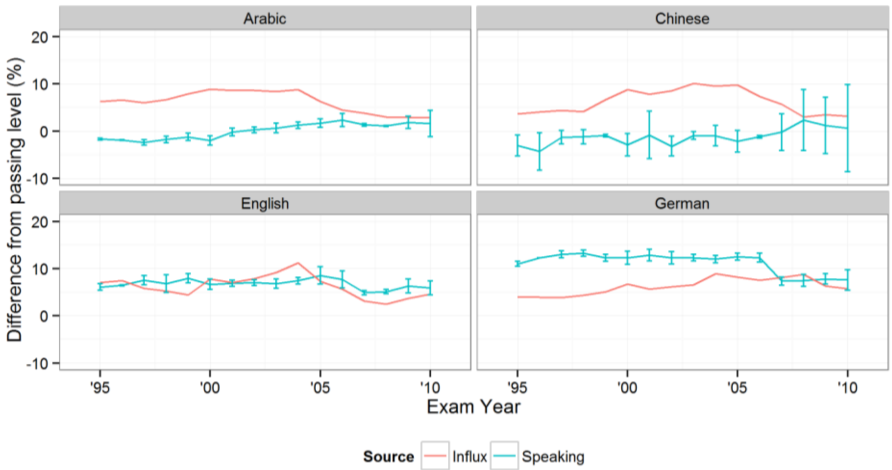


Figure 2. The average differences in speaking proficiency relative to the passing level fluctuate over time and slightly regress towards 0. The scores for German are relatively high. Influx stands for the rate of

immigration per group over all 15 years. The error bars represent standard errors.

Statistical Modeling

Models of linguistic variation have received renewed attention in high profile scientific journals (Atkinson, 2011; Bouckaert et al., 2012), as researchers employ more sophisticated statistical modeling techniques and larger data sets. For example, hierarchical linear regression models are currently a popular tool for the purpose of controlling for genealogical and areal dependencies between languages (Jaeger, Pontillo, & Graff, 2012; Ladd, Roberts, & Dediu, 2015). We investigated how these statistical models can be used to distinguish between individual learner, language, and country level variation in L2 speaking proficiency test scores, and to check the consequences of assuming that these levels vary independently.

There are about 7000 different languages (P. Lewis et al., 2013). In our studies, we made use of learner data that allowed us to compare aggregated proficiency scores across 74 different L1s. These languages are not independent from each other, but belong to different language families that conform to lineage specific trends (Dunn et al., 2011). Such dependencies can lead to overestimations of correlations between characteristics when they are assumed to be independent. This problem can be referred to as an instance of Galton's problem and stands firm in linguistics and other disciplines. (Gavin et al., 2013; Ladd et al., 2015; Levinson & Gray, 2012; Nettle, 2012; Roberts & Winters, 2013). The languages in our data set belong to 14 different language families but many of the L1s are Indo-European (35). To what extent does this particular set of L1s drive the distance effects observed? First, we tried to control for Galton's problem in Chapter 4 by testing whether a general effect of morphological distance still holds after we allow random slopes for morphological distance across language families in linear mixed effect models. Secondly, we let the natural flow of immigration to the Netherlands decide what L1s were included in our study. This is not a random sample of the world's languages, but it means that we did not introduce more bias to the sample (Collier, 1995; Ragin, 2004). However, as linguistic distance, geographical distance, and traveling-time may play a role in immigrants' choice of a host country, it may be the case that the number of distant languages from Dutch is underrepresented in our study, due to self-selection. Evidence is available that shows that immigrants optimize their migration decision depending on the smallest possible linguistic distance amongst others (Adsera & Pytlikova, 2012; Beenstock et al., 2001; Desmet, Ortuño-Ortín, & Wacziarg, 2012; Esser, 2006). We think, however, that the impact of such migration decisions is relatively small for the Netherlands as compared to countries with world languages such as English, French or Spanish. We do not think that relevant immigrant groups have avoided the Netherlands because of a supposed linguistic or cultural incompatibility.

Furthermore, one problem with current mixed effect modeling when applied to partially crossed observational data is collinearity between random factors. We restricted our models to conservative lower bounds estimates of by-L1 country variation by crossing L1s with countries of birth. This crossing allows us to estimate their independent contributions to the overall variance. Removing countries of birth as random effect results in a possible *upper bound* for the degree to which L1 background matters, as this neglects by-country differences. As the model assumes independence between random effects, crossing two random effects helps the model to identify the variation that cannot be attributed to another random effect. This approach shows that crossed models in unbalanced data sets identify more ambiguous variance than simple multilevel models with one random effect only. The comparison between different random effect structures is useful to illustrate the lower and upper bounds of different factors. The difference between the lower and upper bound can be defined as ambiguous. Given the ambiguous variance, we cannot be certain about the precise degree of by-L1 variation. As we can a priori assume that country-related characteristics play a role (e.g. due to factors of post-communist regimes, war history, GDP), we included country as a random factor, although this leads to more conservative by-L1 estimates. The by-L1 variation may thus be larger than we can currently estimate.

With respect to interpreting the distance effects, multilevel models allowed us to account for individual level factors while we tested for language level effects of linguistic distance. At the level of languages and countries, we did not incorporate many variables at the same time, to avoid collinearity problems. Collinearity problems make it hard to disentangle the effects of the different linguistic distance measures. Substantial correlations between different distance measures, as illustrated for phonological and morphological distance in Figure 3, overshadows drawing straightforward conclusions about the individual contributions of linguistic distance effects. We need domain-specific speaking proficiency ratings to get answers to theoretical questions of domain-specificity in distance effects on L2 learnability.

Are there alternative explanations for the effect of linguistic distance? It would be coincidental when geographical or cultural distance is associated with low L2 Dutch learnability in the same way as linguistic distance is (e.g. a low relative complexity compared to Dutch). Such an alternative explanation assumes a correlation between the distribution of linguistic distance (e.g. morphological complexity) and geographical or cultural distance between languages / countries, which would obstruct the possibility to test effects of linguistic distance without controlling for geographical distance. In relation to our measure of relative morphological complexity, our measure is different from absolute measures of morphological complexity as it crucially depends on what features are complex in Dutch and it disregards other features in which a language can be morphologically complex. This does not rule out that geographical and cultural factors do not play a role of course. However, Chapter 6 showed that a simple measure of geographical distance does not explain away the effects of linguistic distance.

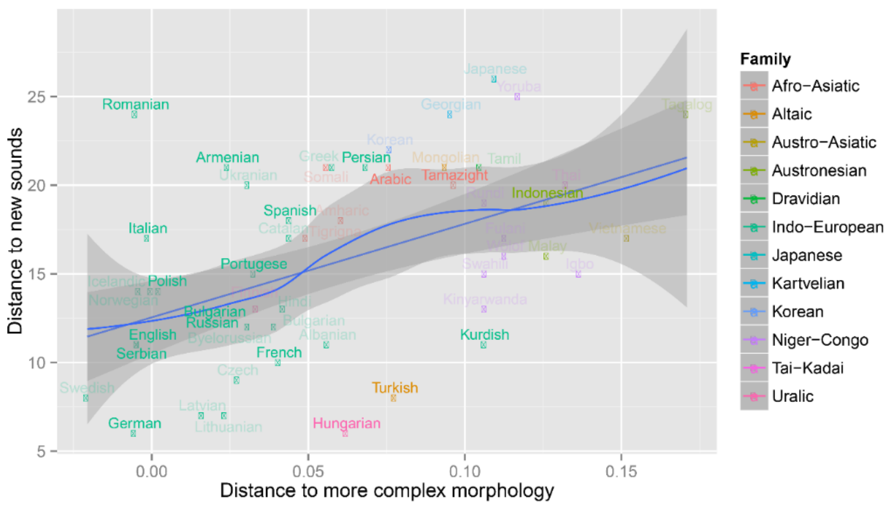


Figure 3. The relation between phonological and morphological distance.

Typological Data

A crucial part of this thesis depends on the availability of typological data. Without the large-scale databases that have recently become available (C. H. Brown et al., 2008; Dryer & Haspelmath, 2011; Gray & Atkinson, 2003; Holman et al., 2008; Moran et al., 2014; Moran & Wright, 2009), the studies reported here would not have been feasible. We linked typological data to data from STEX via the ISO 639 language codes from Ethnologue and WALS. STEX language names were available in Dutch only, so we had to translate and match the Dutch language names to ISO 639 first, which was sometimes ambiguous.

Furthermore, missing data in Gray & Atkinson (2003), ASJP, WALS, or PHOIBLE resulted in a different number of languages for every particular study. Adding data for missing languages was not straightforward. For example, we tried to add data for a number of missing languages in PHOIBLE (e.g. Belarusian, Latvian, Malayalam, Tamazight, Tigrigna, Afrikaans, Papiamentu, Slovak, etc.) by looking up phonological inventories in published grammars. We decided that relevant expertise is necessary to perform this task adequately, which was beyond the scope of this project. Difficult decisions had to be made on which sounds to include, when they only occur with low frequency, in specific positions or contexts, or are borrowed from other languages. As our studies tested distances from a wide variety of languages to Dutch, we always cross-checked the typological data for Dutch. In the future, we hope that alternative typological databases become available that include more faithful phonological transcriptions for computing lexical distances similar to Wieling et al. (2014), more morphological features and feature values for computing morphological distances (e.g. Hammarström & O'Connor, 2013), and suprasegmental information for computing phonological distances.

Importance

We now turn to the importance of our findings for L2 learnability, L3 learnability, learnability constraints, and third factors in additional language learning.

L2 Learnability

Influence of the language background on language learning implies that not all language learners start from the same starting point. We argue here that linguistic distance effects on learning difficulty reveal what linguistic differences lead to the most severe problems in additional language learning. The present thesis employed formalized measures of linguistic distance to study constraints on the learnability of additional language structures. Distance effects show the problems that adult language users have discovering the structure of an additional language. Distances between L1 and L2 structure co-determine the learnability of L2 structure.

We provided insights into the learning algorithm by presenting distance-based models of the degree of difficulty required to learn an additional language. We claim that distances need to take a central role in formal accounts of adult learnability of additional language. Furthermore, because of their prominent role in L2 learnability, distance effects are likely to propagate in language transmission as well. Others scholars have also used the term language learnability to give a formal account of the language-learning algorithm (Apoussidou, 2007; Clark & Roberts, 1993; Jager, 2003; Lightfoot, 1999; Oudeyer, 2001; Perfors, Tenenbaum, & Regier, 2011; Pinker, 1984; Prince & Smolensky, 1997; Tesar & Smolensky, 1996, 1998). An important aim of such (computational) theories of learnability is to explain how faithful language transmission works across generations.

Effects of linguistic distance in learnability are important for L2 learnability theories that aim to identify constraints on linguistic variation. When we generalize our findings on L2 learnability, we also expect that accounts of learnability in general should account for the effect of distance between previously discovered and to be discovered structure. Such an expectation aligns with the finding that general language learning mechanisms gradually discover structure in new input by means of generalization (Solan, Horn, Ruppin, & Edelman, 2003, 2005; Tenenbaum & Griffiths, 2001). In addition, within the optimality theory framework, it has been argued that the learning mechanism crucially depends on the explanatory power of previously encountered structures (Tesar & Smolensky, 1996, 1998). Given our empirical results, we believe that the explanatory power of L1 structures is crucially intertwined with distance effects. The problem of how learners overcome linguistic distances is relevant for understanding the underlying mechanisms of additional language learnability as well as L1 learnability. Indeed, distance effects seem to play a crucial role in general learning and structure discovery problems (Hahn, 2006, 2014; Tenenbaum & Griffiths, 2001; Tversky, 1977). In general, both L2 and L1 learners need to grasp incoming signals by making inferences based on the structure of previously experienced signals.

Distance effects in L2 learnability reveal part of the mechanisms that language learners use to learn additional languages. Formal (procedural) accounts of learnability are crucial in Universal Grammar approaches to language learnability (Gibson & Wexler, 1994; Prince & Smolensky, 1997), in Universal Grammar approaches to L2 learnability

(Archibald, 1994; R. Hawkins, 2001; R. Hawkins & Chan, 1997; Joo, 2003; Slabakova, 2006), and Optimality Theory (Apoussidou, 2007; Tesar & Smolensky, 1996, 1998). Such accounts do not show explicitly that distance effects constrain theories of learnability. Explicit accounts of the role of distances may be crucial to ensure compatibility of these theoretical accounts with large-scale L2 learnability data.

L3 Learnability

Evidence for L2 distance effects in L3 learnability further establishes the idea that learners make use of similarities and differences to make inferences when learning additional languages. Chapters 5 and 6 showed that a number of factors matter: 1) the similarity of the first language, 2) having learned an L2, 3) the similarity of the L2, and 4) the learning order (which language was learned first). Besides the importance for L3 learnability itself, e.g. in terms of the roles of the L1 and the L2 in L3 learnability, these findings also speak to questions about general cognitive benefits of being multilingual (Jessner, 2012, 2014; E. Klein, 1995; Thomas, 1988). It seems that language background can explain differences in L3 learnability beyond a general beneficial effect of being multilingual on learning an additional language. That is, distance effects explain variation in L3 learnability beyond hypothesized effects of greater inhibition and cognitive control for multilingual speakers. It seems that the largest effect on language skills is simply the additive effect of having multiple languages to draw inferences from, besides a smaller effect of being multilingual.[13]

[13] The terminology for first and additional language use varies across the literature. Non-native, second language, L2, and additional language learner are used interchangeably. In some sense, this creates a gap between L1 and L2 speakers who both are legitimate language users of the same language (Bonfiglio, 2010; Kramsch, 2012; O'Rourke & Pujolar, 2013)(Bonfiglio, 2010; Kramsch, 2012)(Bonfiglio, 2010; Kramsch, 2012)(Bonfiglio, 2010; Kramsch, 2012). Similarly, the notion of an L2 only functions to make language level distinctions. At the level of personal language use, the distinctions between L1, L2, L3, and Ln become blurred.

We expect that these differences across learners with different language backgrounds are also apparent within the development of individual learners. We compared L2 and L3 learnability across language backgrounds. We aggregated over time-related individual differences in e.g. tiredness, hours of study, readiness, etc. Not only studies of the developmental trajectory can contribute to accounts of L2 and L3 learnability, we think that comparative studies of the differences in language backgrounds across learners are necessary as well.

Learnability Constraints

Learnability research from the perspective of linguistic typology and language change has argued quite often that universal tendencies in language structure can be explained by learnability constraints (Andersen, 1988; Braunmüller, 1990; Hill, 1978; Trudgill, 1974, 1983), as might even be traced back further to pioneering work on language form to learning relationships (von Humboldt, 1836). Effects of learnability constraints on language change have recently gained renewed attention from evidence from artificial language learning studies (Fedzechkina, Jaeger, & Newport, 2012; Rafferty, Griffiths, & Ettlinger, 2011; Tily, Frank, & Jaeger, 2011). We showed what linguistic differences make learning an L2 difficult. Chapter 3 described the relevance of these differences to learnability constraints. Depending on their language background, learners can make more or less use of existing L1 structures. L2 learners whose L1 is fairly close to the L2 can make relatively profitable use of distance (e.g. as a heuristic) to determine what L1 structures they can transfer and what L1 structure interfere with L2 structures. However, the linguistic differences that generally lead to problems in L2 learnability, independently of the L1, reflect limiting learnability constraints in learning complex morphology. For example, learners generally find it difficult to use verbal person marking when it is not present in the L1.

Linguistic patterns of variation between languages are not uniformly distributed. We know which languages belong to which language families from historical linguistics and we know the structural types to which languages can belong from linguistic typology. Such analyses demand much expertise. Comparative and typological methods, for example, require careful assessment of recurrent sound correspondences and comparisons of grammatical structures (Campbell, 2004; Greenberg, 1973). It may thus seem surprising that language learners draw inferences from their L1s in a manner that reflects genealogical and typological distance, without having to be aware of the distances themselves. After all, the linguistic distance measures used in this study predict variation in proficiency in Dutch in a substantial way. Language learners extract regularities between languages and use them to optimize their learning efforts. Understanding this optimization process may shed light on linguistic distance as well.

To what extent do distributions in typology relate to second language learning? Our argument goes back to the way languages in contact influence each other's structure due to interference phenomena. Following Weinreich (1953), interference occurs when a speaker is exposed to multiple language varieties but produces language that deviates from all of them. Learning a language reveals how the new variety interferes with the learner's language background. Interference in L2 learnability may constrain – ceteris paribus – the spread and growth of languages on speaker populations of other languages. We illustrate this hypothesis with an example.

During the Muslim conquests, classical Arabic (an Afro-Asiatic language) was distributed to speakers from a large geographical area who spoke several different languages that would function as substrates. Signs of these substrate languages can still be observed in the variants of classical Arabic that arose after the Muslim conquests: the variant of Arabic spoken in Syria has an Aramaic substrate, the Egyptian variant has Coptic and Nubian substrates, and the Algerian variant has Berber and Punic substrates. Conversely, there is no variant of Arabic (or Persian) spoken in Anatolia with a Turkic substrate, although Arabic heavily influenced the Ottoman Turkish language and Turkish borrowed extensively from the Arabic and Persian vocabularies.

Perhaps, Arabic and Persian are relatively hard to learn for speakers of Turkish as opposed to speakers of Aramaic, Berber, Punic and Coptic, which are also Afro-Asiatic languages like Arabic (Ostler, 2005). In the conquered areas where Afro-Asiatic languages were spoken, some dissimilar properties of the original varieties were stable enough to introduce innovations into what are now seen as dialects of Arabic. For example, Nubian farming terminology remained persistent in Sudanese Arabic. Many substrate properties were assimilated and similar properties transferred. During the history and formation of Ottoman, however, a relatively large number of properties were either dissimilar or highly stable and may have prevented an overall conversion to an Arabic or Persian variant with an Ottoman substrate. L2 learnability may have played a role in these language contact situations. If these ideas can be confirmed by linguistic analysis, this example is an illustration of how L2 learnability constraints may produce patterns of borrowing and transfer. Distance effects in L2 learnability seem to influence patterns of borrowing and transfer. In contrast to evolutionary constraints, which drive languages apart, patterns of borrowing and transfer bring languages closer to each other.

In an analysis of New York and Chicago dialects, Labov (2007) showed how diffusion across family trees results from language interference, causing unfaithful transmission. In addition, evidence from theoretical approaches in language dynamics (Gong, Shuai, & Zhang, 2014; Mira & Paredes, 2005; Nowak, Komarova, & Niyogi, 2002) further supports the view that distance effects in learnability propagate to language dynamics and change.

It is not completely transparent what distance actually means for language learners, why languages can be distant, why some languages are more distant than other languages, and how distance can change over time. Furthermore, it is not straightforward to implement the intuitive notion of language distance in a rigorous measure. We identified a number of characteristics of distance by comparing distance to L2 learnability. First, L2 learnability is useful for comparing different measures of linguistic distance (Chapter 2). Secondly, L2 learnability is useful for estimating whether learning a more or a less morphologically complex target language is more difficult (Chapter 3). This way, L2 learnability data may contribute to quantitative diachronic investigations of the cultural-evolutionary mechanisms of language variation and change (Nettle, 2012). Transmission of complex morphology in adults seems to be hampered, which is in line with experimental and longitudinal studies of adult L2 learning of morphosyntax (Birdsong & Molis, 2001; Flege, Yeni-Komshian, et al., 1999). Moreover, adult L2 learning depends on the complexity of the L1 of a learner and on the complexity of the target language. L2 learnability can also bring to light a rank order or hierarchy of feature impact, as a measure of accessibility for transfer. Morphological differences low in the hierarchy of impact may be more accessible for recalibration, and features higher up in the hierarchy may be more likely to cause substrate effects. The L2 literature offers evidence that supports this hierarchy. An important problem for L2 learners is to deduce the most optimal constraint ranking (Tesar & Smolensky, 1998). It would be interesting to compare empirically established hierarchies to optimal rankings.

There are no widely accepted models of linguistic distance available for languages stemming from different language families. Morphological and phonological data turned out to be useful for deriving distance measures for languages from different language families. Figure 4 shows the distribution of phonological similarity on a map to illustrate its spread and distribution geographically, based on the languages studied in Chapter 4. Deriving morphological and phonological similarity measures from L2 proficiency scores resulted in asymmetrical measures of linguistic distance (Chapters 3 and 4). In contrast, the evolutionary approaches to linguistic distance that we evaluated define symmetrical relations (Chapter 2). The assumption of symmetry may be unsupported for second language learning and relaxing this constraint revealed what morphological and phonological features are important for second language learning.

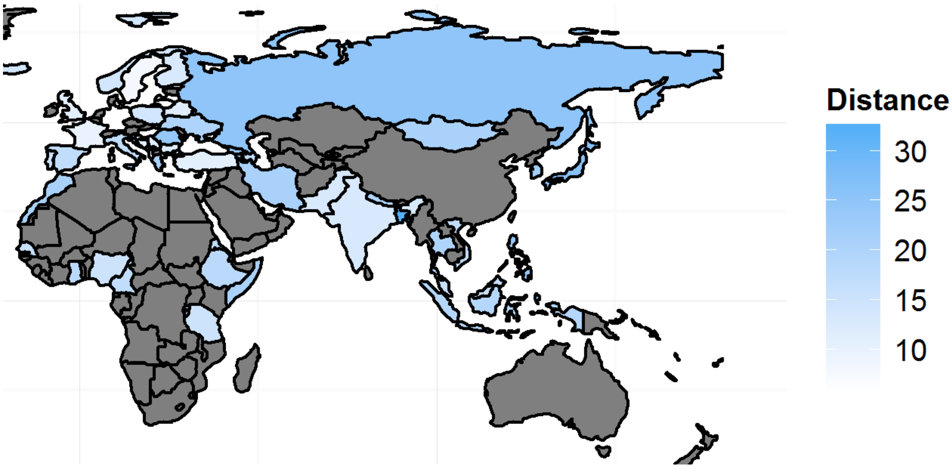


Figure 4. The distribution of phonological distance to Dutch across the languages studied in Chapter 4. For visualization purposes, coloring is removed for countries with missing values and for countries whose main language is primarily spoken elsewhere. Primarily, in this case, means most widely accepted area of origin. We did not have data for languages primarily spoken in the Americas. For the Indian subcontinent, we chose the value for Hindi.

Third Factors in L2 Learnability

Much is known about age effects on second language acquisition (Bongaerts et al., 1997; Flege, Yeni-Komshian, et al., 1999; Johnson & Newport, 1989; Lenneberg, 1967). One common explanation for age-related limitations involves maturational constraints due to reduced plasticity of the adult brain. Indeed, we found that L2 proficiency increases with younger age of acquisition in the learners of L2 Dutch (see for discussion Chapters 2 and 4). The L1-L2 similarity interacted with age and exposure, such that transfer from L1 to L2 a) becomes stronger the later in life the L2 is acquired, and b) weaker with increasing exposure to L2. Although we found that the L1 has a robust influence on L2 learning, the effect is stronger when learners are older (late age at onset of L2 learning) and decreases as experience with L2 increases. Although these interaction effects do not directly falsify a critical period hypothesis as only little data points were available for ages younger than 18, interaction effects provide evidence for experience-based explanations that have also been adopted in other large-scale observational studies (Chiswick & Miller, 2008; Hakuta et al., 2003), meta-analysis (Vanhove, 2013), and L3 research (Pajak et al., 2014). In this sense, a higher age is seen as a container effect representing the effect of more experience as one grows older.

In addition to age and exposure, we controlled for gender and education, which are well-known factors in L2 learning (Bacon & Finnemann, 1992; Boyle, 1987; Labov, 1972; Schmidt, 1977). We found that female learners perform better than male learners do, and that higher educated learners perform better than lower educated learners do. Independently from these effects, linguistic distance has a continuous effect on learning. No interactions of distance with gender and education were observed (see also Van der Slik, Van Hout, & Schepens, in preparation). The advantage of female learners is consistent across many countries and language backgrounds and does not interact with age or exposure. Genetic and hormonal differences between males and females affect brain development, resulting in ability differences as well as different cultural expectations, effectively leading to variation in assessed performance. Such explanations have also been used in studies on assessed performance in mathematical skill (Ceci, Williams, & Barnett, 2009).

As developmental studies comprise an important part of the L2 literature, we were motivated to find out how important age effects are in comparison to language background effects. In STEX, age of arrival and length of residence together explain only a small percentage of the total variance. Language background and country of birth are responsible for about 50% of the residual variation after individual differences and education have been accounted for (see Table 3 in Chapter 5), and other individual differences make up for the remaining variance in L2 learning, which is about 50% as well. Is this variance distribution surprising? We conducted a small-scale questionnaire (n = 89), see Figure 5, to compare general intuitions about the importance of language background, age, exposure, gender, and education for speaking proficiency in adult L2 learning. Almost 40 of the respondents held a PhD, and an additional 27 held a master's degree. Recruitment was conducted in the professional networks of T. F. Jaeger and J. J. Schepens (Jaeger Lab Blog, Twitter, and Facebook), so expertise in linguistics was very common among the respondents. In addition, nearly all of the respondents were at least bilingual. The results depicted in Figure 5 display that respondents estimated effects of language background and country in accordance with our empirical findings, although they seemed to value effects of age and length of residence on adult L2 speaking proficiency more strongly. The importance of age in STEX is relatively low as compared to general intuitions, which suggests that a higher age is not such a great impediment as is often assumed.

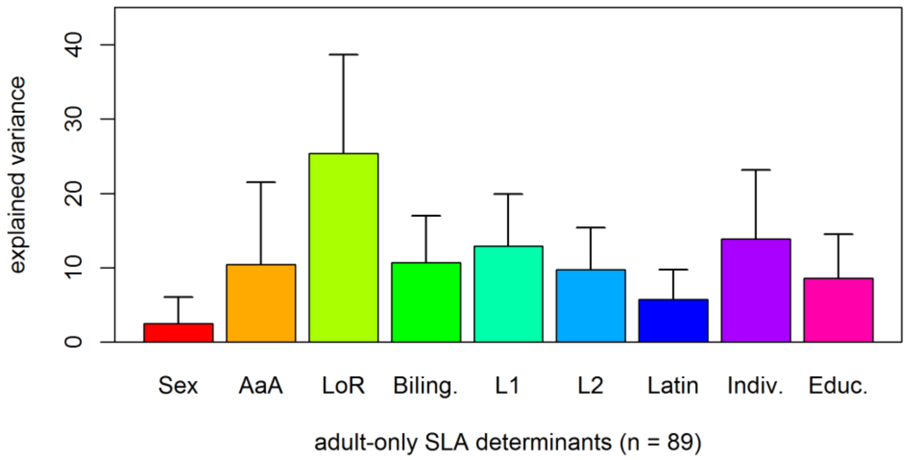


Figure 5. Respondents believe that age at arrival (AaA) and length of residence (LoR) together explain about 35% of the variance in L2 learning, which is a considerably higher percentage than STEX shows. Respondents estimate the role of language background (being bilingual, L1, L2, and Latin alphabet) and education together at about 50%. Individual differences other than age and exposure explain no more than 15% of the variance.

Future Research

There are many reasons to conduct further joint investigations of typological and language testing data. Amon the promising opportunities are: 1) Understand initial and early stages of language development using data on age and exposure to discover at what stages effects of the L1 are most important. 2) Understand modality effects using the possibility to link together re-exams for assessing improvement over time across modalities (reading, listening, writing, and speaking). 3) Understand motivation effects using data of taking the language exam voluntarily or obligatorily. At the same time, it is necessary to look at language testing data for other target languages as well.

A pressing issue that needs further research is the domain specificity of distance effects. As outlined in the Statistical Modeling section above, further disambiguation between the effects of the different distance measures may be important. Domain-specific pronunciation and grammatical proficiency measures are necessary to disentangle the effects of phonological and morphological distance. Future work needs to model proficiency in the specific domains of interest explicitly. A learner corpus study, which can make use of existing data (Cucchiarini, C., Driesen, J., Van Hamme, H., & Sanders, E., 2008; Lüdeling, Walter, Kroymann, & Adolphs, 2005; Nicholls, 2003) is necessary to determine the contribution of phonological and morphological distance. One apparent restriction of learner corpus research so far is the low number of speakers per L1 and the low number of L1s that can be compared.

One pilot study of STEX item level data suggested that data for about 2000 examinees (14 L1s) is not enough to detect distance effects. Item level data contains specific ratings on four point scales for pronunciation and grammar, which is still a highly abstract way to compare across L1s. Ideally, even more L2 specific data may be necessary. Another pilot study using transcriptions of the L2 speech signal from the spoken adult L2 learner Dutch corpus (Cucchiarini, C. et al., 2008) allowed comparison of 18 hours of transcribed speech for 22 L1s, although with often only 1 speaker per language. There is a trade-off between data availability and specificity of L2 proficiency measures.

Another pressing item involves testing of experience-based explanations of learning difficulty. Our current findings suggest that the effects of distance increase with age and decrease with exposure. Further research is needed to determine to what extent these interactions are a result of continued experience, which allows learners to infer the structural characteristics of the L2.

Besides lexical, morphological, and phonological distance measures, future research programs can now aim at determining overall measures of (typological) linguistic distance, e.g. by incorporating further syntactical characteristics like branching direction and word order. The resulting typological measures are useful for researchers who want to explain empirical measures of Dutch L2 proficiency in terms of linguistic differences (Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012). Such future work further aims at developing the linguistic distance measures that are currently available.

Furthermore, future research could aim at making distance measures available for language pairs that do not include Dutch. Understanding the applicability of the distance measured developed here calls for work on other languages besides Dutch. Given that distance measures successfully explain L2 Dutch learnability, it is rational to expect that distance effects explain L2 learnability of other languages, too. Actually, there may be no grounds to assume that these effects are specific to Dutch, although further research of especially non-Western European L2s would be desirable. In addition, testing these measures in other L1, L2, L3 settings would be useful for evaluating to what extent the present calculations can be extrapolated.

Chapter 3 demonstrated that distance measures from L2 learnability research are important for understanding learnability constraints in language structure. The connection between learnability biases and language structure is currently being investigated mainly at a cultural level (Bentz et al., Submitted; Bentz & Winter, 2013; Kortmann & Szmrecsanyi, 2012; Lupyan & Dale, 2010, 2015; McWhorter, 2011; Szmrecsanyi & Kortmann, 2009; Trudgill, 2011). Such studies have provided convincing arguments for variation in the stages of development across languages. Although the cognitive make-up of the speakers of different languages may be comparable, languages vary according to their social structure in which they are embedded. It has long been questioned what cognitive characteristics of speakers propagate into language structure (E. Bates & MacWhinney, 1981; Gibson, 1998, 2000; J. A. Hawkins, 2004; Slobin, 1973). There is, however, still potential for research in making connections between cognitive and cultural levels of investigation (Ladd et al., 2015), especially in view of recent advances in statistical analyses and databases.

Computer simulation may become an important tool to gain insights into the necessary components of the L2 learning procedure. Distance measures have been used in computer simulations of language competition before (Gong et al., 2014; Mira & Paredes, 2005; Nowak et al., 2002; Oudeyer, 2005; Steels, 2011). Such approaches may help to shed light on the learning biases involved in language structure. However, theoretical approaches need to be connected to real-world language use, i.e. meaningful linguistic units and their embedding into context. It has not yet been explored in detail how real linguistic data can be used in connection with theoretical approaches.

Future formalizations from computer simulations may shed light on the way L2 learnability constraints play a role in the optimization of language structure for communicative success. For example, a redundantly complex signal supports robust information transmission in noisy situations. More generally, a redundant signal over-specifies unexpected signals, possibly at the cost of expected signals. It may be that processing constraints forces adult L2 learners to reduce over-specification due to an increased cognitive load in retrieval from memory and competition (Gordon, Hendrick, Johnson, & Lee, 2006; R. L. Lewis, Vasishth, & Van Dyke, 2006; Szmrecsanyi, 2004).

Conclusions

We observed L1 and L2 effects of lexical, morphological, and phonological linguistic distance on the learnability of Dutch as an additional language in adult learners with a variety of language backgrounds. The innovative aspect of this study was the finding that newly developed morphological and phonological measures of linguistic distance successfully explain by-L1 and by-L2 variation in language testing scores of more than 50,000 adult learners of Dutch. This brings together different perspectives in linguistics and cognitive science to result in a novel perspective that looks at L2 learning and linguistic differences simultaneously. Apart from existing studies that revealed distance effects in testing scores from applied economics (Isphording & Otten, 2013; Kim & Lee, 2010) and SLA (Van der Slik, 2010), this study demonstrates distance effects for newly developed and more specific distance measures across both L1s and L2s. We gave specific accounts of morphological and phonological transfer that show what differences result in higher learning difficulties. We conclude that linguistic distance influences L2 learnability by constraining the ability to use similarities and differences between the language background and the target language. This conclusion challenges views that place strong emphasis on interference effects between closely related languages. In addition, we conclude that the role of distance in L2 learnability depends on the specific processing constraints of the learners. In adult L2 speakers, the ability to learn additional morphologically complex constructions seems to have diminished. A measure of distance based on morphological complexity provided evidence at the level of adult L2 learning for adaptation effects in language structure (Lupyan & Dale, 2010; Trudgill, 2011). With respect to learning new sounds, we conclude that the learning load posed by different new sounds, which require a learner to expand his or her internal feature geometry, is higher than the learning load of similar new sounds. The expansion of internal feature geometries constrains L2 learnability more than the learning load that is introduced by similar new sounds.

Our resulting model of L2 learnability assigns a pivotal role to distance effects. The theoretically motivated algorithms that we developed for computationally determining morphological and phonological distances present new formal models of linguistic distance. These formal models confirm the importance of morphological and phonological differences for adult L2 learners. These formalized measures of morphological and phonological distance are consistent with predictions of existing lexical distance measures. At the same time, they go beyond lexical distance measures, as they can measure distances between languages from different language families.

Distance effects may play a more important role in L2 learnability as compared to L1 learnability. However, the function of distance may be elementary to the general learning mechanisms that are available to intelligent systems capable of generalization and adaptation to new environments. Such an underlying explanation would be able to account for the observed distance effects in L2 learnability. Further research is needed to figure out the domain-specificity and the experience-based nature of the distance effects in L2 learnability. Such research can be conducted with further large-scale quantitative simulation and learner corpus research. Future studies should also give distance a pivotal role in theories of learning mechanisms underlying learnability in language learners.

The studies in this dissertation resulted in research products in the form of usable and validated linguistic distance measures, see Appendix A. The distance measures are validated on empirical L2 speaking proficiency scores, controlling for age, exposure, education, and gender. The distance measures may be useful for researchers who want to control for linguistic distance in measures of L2 performance. The formalized distance measures are applicable in computational tools for automated language analysis, manual and machine translation, and computer assisted language learning to produce predictions of learnability across Indo-European as well as non-Indo-European languages. With respect to societal benefits, acknowledgement of linguistic gaps may be an important step for development of an evidence-based language learning strategy.

Appendix A

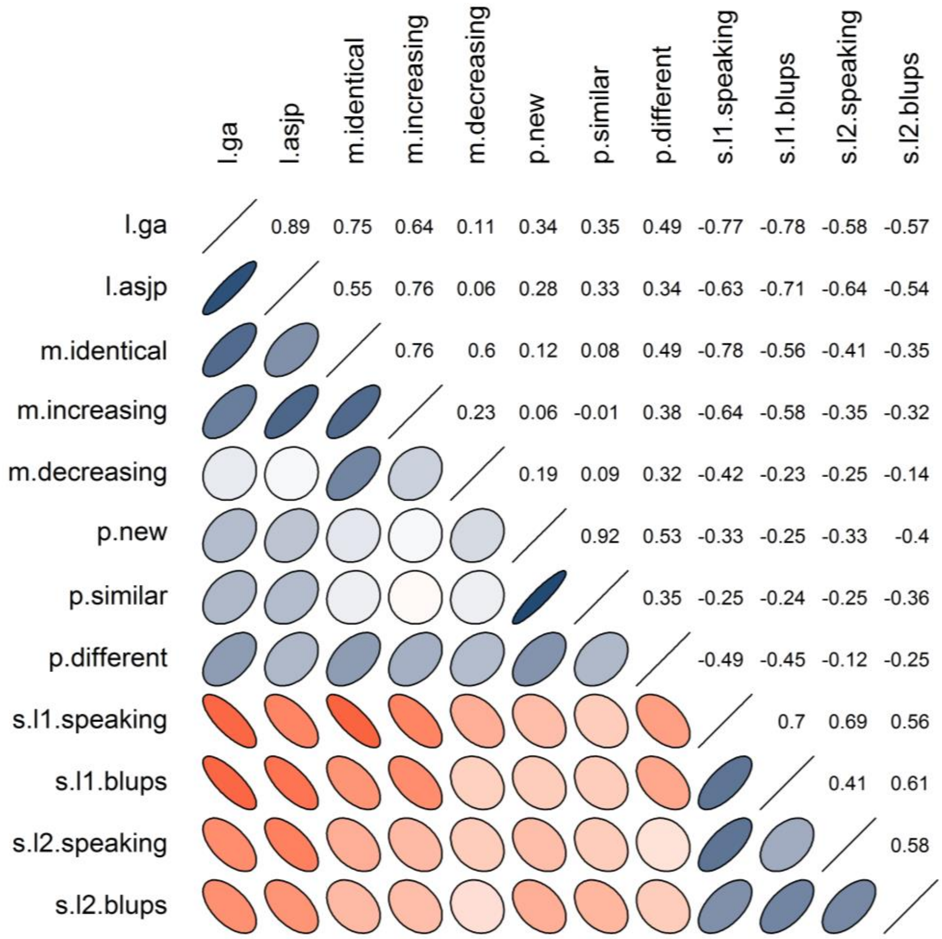


Figure 1. Correlation matrix describing the relation between the distance measures described in Chapters 2-4 and L1 and L2 speaking proficiency.

Note. The variable names represent two measures of lexical distance (l.ga, l.asjp) as based on cognacy judgments (Gray & Atkinson, 2003) and automatic similarity judgment (Brown et al., 2008) from Chapter 2, morphological distance (m.identical), increasing morphological complexity (m.increasing), and decreasing morphological complexity (m.decreasing) from Chapter 3 as based on the World Atlas of Language Structures (Dryer & Haspelmath, 2011), new sounds (p.new), shared features with new sounds (p.similar), complementary features of new sounds (p.different) as based on the Phonetic Information Base and Lexicon (Moran, McCloy, & Wright, 2014), speaking proficiency aggregated by L1s (s.l1.speaking) and controlled for age, exposure, gender, and education (s.l1.blups), the number of learners per L1, speaking proficiency aggregated by L2s (s.l2.speaking) and controlled for age, exposure, gender, and education (s.l2.blups), the number of learners per L2, and the WALS ISO language identification code

Table 1. Measurements of linguistic distance and L1 and L2 speaking proficiency for the languages studied in this thesis. Variable names are described in Figure 1.

l1	l. ga	m. incre asing	p. diff	s.l1. speak ing	s.l1. blups	s.l1. n	s.l2. speaki ng	s.l2. blups	s.l2. length	wals iso
Afrikaans	-1.78			544.09	18.66	301	541.98	1.51	130	afr
Albanian	1.56	-.16	-.61	510.52	-5.53	288	511.81	-0.6	16	als
Amharic		.27	.63	490.36	-3.34	138				amh
Arabic		.24	1.16	498.93	-3.88	5927	496.73	-3.41	1315	ary
Armenian	.98	-.36	1.16	512.16	7.11	625	513.91	2.14	68	hye
Azerbaijani			.45	510.79	0.59	129	504.16	-1.34	25	azj
Bengali	1.16		3.11	494.69	-7.37	42				ben
Bosnian		.24	-.61	527.47	4.96	99				hrv
Bulgarian	.51	.25	-.44	528.14	4.22	563	495.42	-2.21	31	bul
Belarusian	.44	-.06	-.44	533.73	9.23	26				rus
Catalan	.42	-.55	.45	524.87	-2.95	70	521.74	0.37	43	cat
Chinese		4.31		492.89		913	511.64	-1.77	64	cmn
Croatian	.42	.24	-.61	516.4	-0.7	481	502.59	-2.73	27	hrv

Czech	.36	.07	-.97	540.05	21.19	471	518.61	2.45	71	ces
Danish	-1.38	-.89		549.73		209	521.92	1.28	36	dan
English	-1.36	-.72	-.61	534.56	17.56	3041	522.57	2.58	27886	eng
Estonian		.08		540.95		65				est
Éwé			.63	496.35	-2.21	26				ewe
Finnish		-.4	-.26	538.05	1.62	238	544.58	-0.18	26	fin
French	.44	-.66	-.79	524.82	2.1	1574	507.03	0.48	4804	fra
Fulani		1.36	.45	499.52		23				fub
Georgian		-.56	1.69	504.69		117	517.04	-0.74	25	kat
German	-1.62	-1.19	-1.50	554.77	30.52	5226	531.65	10.23	2003	deu
Greek	1	-.17	1.16	528.89	3.09	333	522.77	0.22	39	ell
Haitian				493.81	-4.61	27				hat
Hebrew			1.87	523.66	-8.66	261	525.22	1.97	50	heb
Hindi	1.19	-.06	-.26	501.58	-7.66	100	525.45	1.67	40	hin
Hungarian		-.71	-1.50	537.36	7.34	794	525.32	0.35	38	hun
Icelandic	-1.17	-.78	-.08	534.27		41				isl
Igbo		4.89	.09	486.12	-3.8	66				ibo
Indonesian		1.7	.80	501.62	-8.25	1422	504.72	-0.27	32	ind
Italian	.21	-.91	.45	530.72	6	662	513.76	-4.1	348	ita
Japanese		2.21	2.04	499.49	-18.18	278	508.19	-0.49	37	jpn
Javanese			-.44	497.81	0.91	26				jav
Kinyarw.		.21	-.26	497.01	-0.72	68				kin
Korean		1.64	1.33	497.7	-10.31	64				kor
Kurdish		.04	-.61	495.82	-2.58	1127	497.96	-0.93	77	kmr
Latvian	.24	.09	-1.32	527.09	-3.39	75				lit
Lithuanian	.28	.69	-1.32	523.01	-2.88	192				lit
Malay		1.7	.27	506.42	-4.48	50	528.04	-0.09	23	zsm
Malayalam				519.57		21				mal
Mongolian		2.15	1.16	493.14	-10.98	44				khk
Nepali	1.05		.98	495.11		36				nep
Norwegian		-1.02	-.08	549.34		204	540.23	2.03	22	nob
Papiamen.				533.5	13.75	145				pap
Pashto			.63	500.02	-0.1	277	500.85	1.06	40	pst
Persian	1.33	.18	1.16	499.95	-2	3382	499.12	-3.18	396	pes
Polish	.38	-.04	-.08	526.79	2.87	2651	519.8	-2.27	69	pol
Portugese	.5	.15	.09	514.89	0.88	1255	516.43	-0.34	88	por
Romanian	.52	-.4	1.69	526.42	0.03	990	527.19	0.61	84	ron

Rundi		.21	.80	493.31		158				run
Russian	.42	-.06	-.44	520.47	2.14	3772	510.34	-2.28	1889	rus
Serbian	.42	.24	-.61	514.41	-0.81	2188	513.59	-0.05	92	hrv
Slovak	.34			534.47	1.57	335	532.54	2.06	24	slk
Slovenian	.43		.45	549.55	0.77	56				slv
Somali		-.14	1.16	482.46	-8.82	423				som
Spanish	.44	-.55	.63	514.1	-3.15	2872	524.69	-0.46	725	spa
Swahili		.21	.09	496.96	5.41	75	511.52	0.19	23	swh
Swedish	-1.26	-.89	-1.15	552.29	13.05	312	536.56	4.45	70	swe
Tagalog		3.4	1.69	499.74	-5.81	347				tgl
Tamazight		.6	.98	497.7	7.47	602	509	-1.27	30	shi
Tamil		-.08	1.16	517.25	7.23	28				tam
Thai		3.36	.98	488.05	-14.69	265				tha
Tigrig		-.08	.45	494.83	2.52	60				tig
Turkish		0	-1.15	500.42	-16.03	2868	499.39	-4.31	329	tur
Ukranian	.41	-.06	.98	517.67	-0.05	343	515.95	-0.23	132	ukr
Urdu			-.26	505.83	4.48	127	513.23	1.02	13	hin
Vietmese		4.17	.45	490.85	-17.15	182				vie
Wolof		1.36	.27	501.76		17				wol
Yoruba		2.94	1.87	488.96		23				yor

Appendix B

Table 1. Mean speaking scores for the cross-classification of all 35 Indo –European languages by the 89 countries of origin. The table also contains a measure of schooling quality based on the gross enrolment rate in secondary schools (UNESCO, 2011), ASJP linguistic distance measurements (based on Brown et al., 2008), G&A linguistic distance measurements (Gray & Atkinson, 2003), and a count of the number of individuals in each cross-classification.

Country of Birth	Mother Tongue	Speaking	Schooling Quality	ASJP Distance	G&A Character	Group Size
	σ	20	51	1320	0.16	602
	$\bar{x}$	522	448	8442	0.33	278
South Africa	African	551	466	3458	0.0105	265
Albania	Albanian	512	384	9355	0.5951	111
Yugoslavia	Albanian	505	384	9355	0.5951	162
Armenia	Armenian	505	458	9462	0.4930	109
Azerbaijan	Armenian	509	389	9462	0.4930	42
Iran	Armenian	508	440	9462	0.4930	45
Iraq	Armenian	501	391	9462	0.4930	71
Syria	Armenian	501	423	9462	0.4930	31
Turkey	Armenian	517	454	9462	0.4930	19
USSR	Armenian	512	458	9462	0.4930	265
Bosnia	Bosnian	518	462	9107	0.4109	76
Bulgaria	Bulgarian	527	432	9104	0.4111	557
Spain	Catalan	525	484	8880	0.3955	65
Croatia	Croatian	543	474	9108	0.3950	58
Yugoslavia	Croatian	515	474	9108	0.3950	382
Czech Rep.	Czech	541	461	9187	0.3852	353
Czechoslovakia	Czech	534	490	9187	0.3852	104

Denmark	Danish	552	499	6862	0.0808	192
Aruba	English	532	477	6586	0.0832	19
Australia	English	541	519	6586	0.0832	174
Cameroon	English	493	359	6586	0.0832	49
Canada	English	543	527	6586	0.0832	173
Germany	English	550	510	6586	0.0832	32
Ghana	English	498	389	6586	0.0832	29
Guyana	English	512	479	6586	0.0832	29
India	English	527	402	6586	0.0832	47
Ireland	English	541	497	6586	0.0832	155
Liberia	English	485	375	6586	0.0832	37
Malaysia	English	534	427	6586	0.0832	24
Netherlands	English	555	519	6586	0.0832	80
New Zealand	English	538	524	6586	0.0832	76
Nigeria	English	499	371	6586	0.0832	71
Philippines	English	525	449	6586	0.0832	22
Sierra Leone	English	499	359	6586	0.0832	20
Singapore	English	539	543	6586	0.0832	39
South Africa	English	541	466	6586	0.0832	219
United Kingdom	English	540	500	6586	0.0832	983
United States	English	539	496	6586	0.0832	795
Zimbabwe	English	537	375	6586	0.0832	24
Algeria	French	520	449	9012	0.3981	32
Belgium	French	533	509	9012	0.3981	98
Burundi	French	497	336	9012	0.3981	23
Cameroon	French	493	359	9012	0.3981	59
Canada	French	530	527	9012	0.3981	40
Congo, Dem. Rep.	French	491	350	9012	0.3981	65
Congo, Rep.	French	495	384	9012	0.3981	61
Cote d'Ivoire	French	496	357	9012	0.3981	35

France	French	531	497	9012	0.3981	936
Morocco	French	514	394	9012	0.3981	42
Netherlands	French	542	519	9012	0.3981	27
Switzerland	French	550	517	9012	0.3981	37
Austria	German	566	487	5664	0.0373	232
Germany	German	558	510	5664	0.0373	4434
Netherlands	German	560	519	5664	0.0373	25
Poland	German	550	501	5664	0.0373	18
Switzerland	German	571	517	5664	0.0373	190
Germany	Greek	544	510	9440	0.4974	20
Greece	Greek	529	473	9440	0.4974	271
India	Hindi	507	402	9211	0.5302	92
Suriname	Hindi	530	430	9211	0.5302	38
Iceland	Icelandic	539	501	6995	0.1174	36
Italy	Italian	533	486	8858	0.3586	594
Netherlands	Italian	547	519	8858	0.3586	20
Iran	Kurdish	491	440	9241	0.4257	91
Iraq	Kurdish	492	391	9241	0.4257	738
Syria	Kurdish	487	423	9241	0.4257	63
Turkey	Kurdish	490	454	9241	0.4257	185
Latvia	Latvian	546	487	9417	0.3636	28
USSR	Latvian	525	487	9417	0.3636	39
Lithuania	Lithuanian	520	479	9318	0.3711	77
USSR	Lithuanian	523	479	9318	0.3711	113
Nepal	Nepali	490	389	9592	0.5054	18
Norway	Norwegian	555	500	6843	0.1598	175
Afghanistan	Pashto	498	346	9539	0.4588	274
Afghanistan	Persian	495	346	9322	0.5553	1252
Iran	Persian	497	440	9322	0.5553	2063
Poland	Polish	526	501	9313	0.3880	2608
Angola	Portuguese	501	342	8967	0.4087	114
Brazil	Portuguese	514	401	8967	0.4087	784
Cape Verde	Portuguese	503	425	8967	0.4087	72
Mozambiqu	Portuguese	521	336	8967	0.4087	26

Portugal	Portuguese	524	490	8967	0.4087	216
Moldova	Romanian	523	458	8893	0.4137	21
Romania	Romanian	525	426	8893	0.4137	929
Azerbaijan	Russian	512	389	9413	0.3961	40
Belarusian	Russian	530	471	9413	0.3961	83
Germany	Russian	527	510	9413	0.3961	22
Kazakhstan	Russian	515	398	9413	0.3961	42
Russia	Russian	524	468	9413	0.3961	767
Ukraine	Russian	520	464	9413	0.3961	210
USSR	Russian	520	468	9413	0.3961	2521
Uzbekistan	Russian	517	477	9413	0.3961	26
Bosnia	Serbian	521	462	9094	0.3950	28
Serbia	Serbian	524	442	9094	0.3950	98
Yugoslavia	Serbian	513	442	9094	0.3950	2000
Sri Lanka	Singhalese	495	456	9440	0.5393	37
Czech Rep.	Slovak	537	461	9205	0.3810	220
Slovakia	Slovak	529	488	9205	0.3810	95
Yugoslavia	Slovenian	550	499	9009	0.3971	38
Argentina	Spanish	525	396	9117	0.3986	188
Bolivia	Spanish	505	458	9117	0.3986	55
Chile	Spanish	507	439	9117	0.3986	113
Colombia	Spanish	508	399	9117	0.3986	399
Costa Rica	Spanish	528	443	9117	0.3986	37
Cuba	Spanish	507	466	9117	0.3986	100
Dominican Rep.	Spanish	499	425	9117	0.3986	64
Ecuador	Spanish	510	420	9117	0.3986	113
Guatemala	Spanish	518	397	9117	0.3986	38
Mexico	Spanish	513	420	9117	0.3986	293
Nicaragua	Spanish	504	423	9117	0.3986	34
Peru	Spanish	507	368	9117	0.3986	333
Spain	Spanish	524	484	9117	0.3986	799
Uruguay	Spanish	540	427	9117	0.3986	33
Venezuela	Spanish	508	435	9117	0.3986	190

Finland	Swedish	558	544	6890	0.1005	27
Sweden	Swedish	556	495	6890	0.1005	272
Ukraine	Ukrainian	518	464	9387	0.3941	141
USSR	Ukrainian	519	464	9387	0.3941	186
Pakistan	Urdu	499	362	9128	0.4132	109

References

Adsera, A., & Pytlikova, M. (2012). *The role of language in shaping international migration* (SSRN Scholarly Paper No. ID 2003666). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2003666

Ahukanna, J. G. W., Lund, N. J., & Gentile, J. R. (1981). Inter- and intra-lingual interference effects in learning a third language. *The Modern Language Journal*, *65*(3), 281–287. doi:10.1111/j.1540-4781.1981.tb00985.x

Aikhenvald, A. Y., & Dixon, R. M. W. (2006). *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press.

Amaro, J. C., Flynn, S., & Rothman, J. (2012). *Third language acquisition in adulthood*. Amsterdam: John Benjamins Publishing.

Andersen, H. (1988). Center and periphery: Adoption, diffusion and spread. In J. Fisiak (Ed.), *Historical dialectology: Regional and social* (pp. 39–83). Berlin: Walter de Gruyter.

Apoussidou, D. (2007). *The learnability of metrical phonology*. LOT. Retrieved from http://dare.uva.nl/document/41607

Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research*, *10*(3), 215–240.

Ard, J., & Homburg, T. (1983). Verification of language transfer. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 157–176). MA: Newbury House Publishers.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, *332*(6027), 346–349. doi:10.1126/science.1199295

Baaij, C. J. W. (2012). The EU policy on institutional multilingualism: Between principles and practicality. *Language & Law*, *1*(1). Retrieved from http://www.languageandlaw.de/volume-1/3338

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items.

Journal of Memory and Language, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Bacon, S., & Finnemann, M. D. (1992). Sex differences in self-reported beliefs about foreign-language learning and authentic oral and written input. *Language Learning*, *42*(4), 471–495.

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*(3), 339–362. doi:10.1016/j.jml.2004.12.003

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Second Language Research*, *23*(4), 459–484. doi:10.1177/0267658307080557

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv:1406.5823 [stat]; ArXiv E-Print; Submitted to Journal of Statistical Software*. Retrieved from http://arxiv.org/abs/1406.5823

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes (Version R package version 0.999999-0). Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Bates, E., & MacWhinney, B. (1981). Second-language acquisition from a functionalist perspective: pragmatic, semantic, and perceptual strategies. *Annals of the New York Academy of Sciences*, *379*(1), 190–214.

Beenstock, M., Chiswick, B. R., & Repetto, G. L. (2001). The effect of linguistic distance and country of origin on immigrant language skills: application to israel. *International Migration*, *39*(3), 33–60. doi:10.1111/1468-2435.00155

Bentz, C., Verkerk, A., Kiela, D., Hill, F., & Buttery, T. (Submitted). Adaptive languages: Modeling the co-evolution of population structure and lexical diversity.

Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, *3*(1), 1–27. doi:10.1163/22105832-13030105

Beretvas, S. N. (2011). Cross-classified and multiple membership models. In J. J. Hox & K. Roberts (Eds.), *The handbook of advanced multilevel analysis* (pp. 313–334). New York: Routledge.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–206). York Timonium, MD: York press. Retrieved from http://ci.nii.ac.jp/naid/10018033931/

Bialystok, E. (2013). The impact of bilingualism on language and literacy development. In T. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (p. 624). Oxford: Wiley-Blackwell.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*(s1), 9–49. doi:10.1111/j.1467-9922.2006.00353.x

Birdsong, D. (2014). The critical period hypothesis for second language acquisition: Tailoring the coat of many colors. In *Essential topics in applied linguistics and multilingualism* (pp. 43–50). Heidelberg: Springer.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249. doi:10.1006/jmla.2000.2750

Bleakley, H., & Chin, A. (2010). Age at arrival, english proficiency, and social assimilation among us immigrants. *American Economic Journal: Applied Economics*, *2*(1), 165–192. doi:10.1257/app.2.1.165

Blom, E., Polišenská, D., & Weerman, F. (2006). Effects of age on the acquisition of agreement inflection. *Morphology*, *16*(2), 313–336. doi:10.1007/s11525-007-9110-1

Board of Examinations. (2011). *Evaluator profile l2 dutch*. *Dutch as a Second Language*. Retrieved December 11, 2013, from https://www.cve.nl/document/beoordelaarsprofiel

Bohnacker, U. (2006). When swedes begin to learn german: from V2 to V2. *Second Language Research*, *22*(4), 443–486. doi:10.1191/0267658306sr275oa

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and

evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. doi:10.1016/j.tree.2008.10.008
Boll-Avetisyan, N. (2012). Phonotactics and its acquisition, representation, and use: An experimental-phonological study. *LOT*, *298*. Retrieved from http://dspace.library.uu.nl/handle/1874/241655
Boll-Avetisyan, N., & Kager, R. (2013). OCP-PLACE in speech segmentation. *Language and Speech*, 0023830913508074. doi:10.1177/0023830913508074
Bonfiglio, T. P. (2010). *Mother tongues and nations: The invention of the native speaker* (1st ed.). Berlin: De Gruyter Mouton.
Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Erlbaum.
Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*(04), 447–465.
Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., … Atkinson, Q. D. (2012). Mapping the origins and expansion of the indo-european language family. *Science*, *337*(6097), 957–960. doi:10.1126/science.1219669
Boyle, J. P. (1987). Sex differences in listening vocabulary. *Language Learning*, *37*(2), 273–284.
Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. doi:10.1016/j.cognition.2007.04.005
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.
Braunmüller, K. (1990). Komplexe flexionssysteme - (k)ein problem für die natürlichkeitstheorie? *Zeitschrift Für Die Phonetische Sprachwissenschaft Und Kommunikationsforschung*, *43*(5), 625–635.
Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture:

A study of Japanese speakers of English. *Studies in Second Language Acquisition*, *30*(02), 225–251.
Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136–193. doi:10.1191/026765898669508401
Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. *Second Language Acquisition and Linguistic Theory*, 4–63.
Brown, C. H., Holman, E. W., Wichmann, S., & Velupillai, V. (2008). Automated classification of the world′ s languages: a description of the method and preliminary results. *STUF-Language Typology and Universals*, *61*(4), 285–308.
Campbell, L. (2004). *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.
Cavalli-Sforza, L. L. (1997). Genes, peoples and languages. *PNAS*, *94*(15), 7719–7724.
Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton: Princeton University Press.
Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: sociocultural and biological considerations. *Psychological Bulletin*, *135*(2), 218.
Cenoz, J. (2001). The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 8–20). Clevedon, UK: Multilingual Matters.
Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, *7*(1), 71–87.
Cenoz, J., Hufeisen, B., & Jessner, U. (Eds.). (2001). *Cross-linguistic influence in third language aquisition: Psycholinguistic perspectives*. Clevedon, UK: Multilingual Matters.
Cenoz, J., & Valencia, J. F. (1994). Additive trilingualism: Evidence from the Basque Country. *Applied Psycholinguistics*, *15*(02), 195–207. doi:10.1017/S0142716400005324
Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production.

Journal of Phonetics, *40*(2), 249–268. doi:10.1016/j.wocn.2011.10.007

Charette, M., & Meng, R. (1994). Explaining language proficiency: Objective versus self-assessed measures of literacy. *Economics Letters*, *44*(3), 313–321.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*(7), 1131–1157. doi:10.1111/j.1551-6709.2009.01049.x

Chiswick, B. R., & Miller, P. W. (1994). Language and labor supply: The role of gender among immigrants in Australia. *Research on Economic Inequality*, *5*, 153–189.

Chiswick, B. R., & Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, *38*(3), 391–409. doi:10.1353/dem.2001.0025

Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, *26*(1), 1–11. doi:10.1080/14790710508668395

Chiswick, B. R., & Miller, P. W. (2007). Modeling immigrants' language skills. *Research in Labor Economics*, *27*, 75–128. doi:10.1016/S0147-9121(07)00003-9

Chiswick, B. R., & Miller, P. W. (2008). A test of the critical period hypothesis for language learning. *Journal of Multilingual and Multicultural Development*, *29*(1), 16–29. doi:10.2167/jmmd555.0

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row. Retrieved from http://eric.ed.gov/?id=ED020511

Clark, R., & Roberts, I. (1993). A computational model of language learnability and language change. *Linguistic Inquiry*, 299–345.

Clements, G. N. (2009). The role of features in phonological inventories. *Contemporary Views on Architecture and Representations in Phonological Theory*, 19–68.

Collier, D. (1995). Translating quantitative methods for qualitative researchers: The case of selection bias. *American Political Science Review*, *89*(02), 461–466.

Comrie, B. (1984). Why linguists need language acquirers. In W. E. Rutherford (Ed.), *Language Universals and Second Language*

Acquisition (pp. 11–29). Amsterdam: John Benjamins Publishing.

Cook, V. (2013). *Second language learning and language teaching*. Routledge.

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. doi: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. doi:10.1016/j.cognition.2006.12.013

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Crothers, J. H., Lorentz, J. P., Sherman, D. A., & Vihman, M. M. (1979). *Handbook of phonological data from a sample of the world's languages: A report of the stanford phonology archive*.

Crystal, D. (1987). *The encyclopedia of language*. Cambridge: Cambridge University Press.

Cucchiarini, C., Driesen, J., Van Hamme, H., & Sanders, E. (2008). Recording speech of children, non-natives and elderly people for hlt applications: The JASMIN-CGN corpus. *Proceedings of LREC 2008, Marrakech, Morroco, May 26 - June 1, 2008*, CD–ROMnull.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cysouw, M. (2013). Predicting language-learning difficulty. In *Approaches to measuring linguistic differences*. Berlin: De Gruyter Mouton. Retrieved from http://www.degruyter.com/view/books/9783110305258/9783110305258.57/9783110305258.57.xml

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins Publishing.

Dahl, Ö., & Velupillai, V. (2013). The past tense. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/66

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, *15*(3-4).

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: Murray.

Dauenhauer, N. M., & Dauenhauer, R. (1998). Technical, emotional, and ideological issues in reversing language shift: Examples from Southeast Alaska. In L. A. Grenoble & L. J. Whaley (Eds.), *Endangered languages: Current issues and future prospects* (pp. 57–98). Cambridge: Cambridge University Press.

De Angelis, G. (2005). Interlanguage transfer of function words. *Language Learning*, *55*(3), 379–414.

De Angelis, G., & Dewaele, J.-M. (2011). *New trends in crosslinguistic influence and multilingualism research* (Vol. 60). Clevedon, UK: Multilingual Matters.

De Angelis, G., & Selinker, L. (2001). Interlanguage transfer and competing linguistic systems in the multilingual mind. *Bilingual Education and Bilingualism*, 42–58.

Dediu, D., & Ladd, R. D. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *PNAS*, *104*(26), 10944–10949. doi:10.1073/pnas.0610848104

Dediu, D., & Levinson, S. C. (2012). Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLoS ONE*, *7*(9), e45198. doi:10.1371/journal.pone.0045198

DeKeyser, R. (2012). Interactions Between Individual Differences, Treatments, and Structures in SLA. *Language Learning*, *62*, 189–200. doi:10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189–200. doi:10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. M., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(03), 413–438.

Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development*

Economics, *97*(2), 322–338. doi:10.1016/j.jdeveco.2011.02.003

Desmet, K., Weber, S., & Ortuño-Ortín, I. (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association*, *7*(6), 1291–1318. doi:10.1162/JEEA.2009.7.6.1291

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, R. H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284–301.

Dixon, R. M. W. (1997). *The rise and fall of languages*. Cambridge: Cambridge University Press.

Dryer, M. S., & Haspelmath, M. (Eds.). (2011). *The world atlas of language structures online*. Max Planck Digital Library. Retrieved from http://wals.info/

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79–82. doi:10.1038/nature09923

Dunn, M., Terrill, A., Reesink, G., Foley, R. A., & Levinson, S. C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science*, *309*(5743), 2072–2075.

Dyen, I., Kruskal, J. B., & Black, P. (1992). An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society*, *82*(5), iii–132.

EC. (2001). *Special Eurobarometer 55.1: Globalization and humanitarian aid*. European Commission.

EC. (2006). *Special Eurobarometer 243: Europeans and their languages*. European Commission.

EC. (2012). *Special Eurobarometer 386: Europeans and their languages*. European Commission.

Eckman, F. R., Elreyes, A., & Iverson, G. K. (2003). Some principles of second language phonology. *Second Language Research*, *19*(3), 169–208. doi:10.1191/0267658303sr2190a

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(02), 305–352.

Espenshade, T. J., & Fu, H. (1997). An Analysis of English-Language Proficiency among U.S. Immigrants. *American Sociological Review*, *62*(2), 288–305. doi:10.2307/2657305

Esser, H. (2006). *Migration, language and integration*. WZB. Retrieved from http://193.174.6.11/alt/aki/files/aki_research_review_4.pdf

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(05), 429. doi:10.1017/S0140525X0999094X

Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research*, *27*(1), 59–82.

Falk, Y., Lindqvist, C., & Bardel, C. (2013). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism: Language and Cognition*, *FirstView*, 1–9. doi:10.1017/S1366728913000552

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203–12208. doi:10.1073/pnas.0602173103

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, *8*(2), 195–222. doi:10.1023/A:1024419522867

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897–17902. doi:10.1073/pnas.1215776109

Finnie, R., & Meng, R. (2005). Literacy and labour market outcomes: self-assessment versus test score measures. *Applied Economics*, *37*(17), 1935–1951.

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.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589–1608. doi:10.1121/1.406818

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, 319–355.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987.

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. doi:10.1121/1.413041

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English/I/and/l/accurately. *Language and Speech*, *38*(1), 25–55.

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. doi:10.1006/jmla.1999.2638

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, *1*(1), 3–16. doi:10.1080/14790710408668175

Foley, C., & Flynn, S. (2013). The role of the native language. In J. R. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition*.

Forster, P., & Renfrew, C. (2006). *Phylogenetic methods and the prehistory of languages*. McDonald Institute for Archaeological Research.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, *22*(1), 179–228. doi:10.1023/B:NALA.0000005557.78535.3c

Gardner, R. C., & Lambert, W. E. (1972). Attitudes and motivation in second-language learning.

Gass, S. M., & Selinker, L. (1992). *Language transfer in language learning: revised edition*. John Benjamins Publishing.

Gavin, M. C., Botero, C. A., Bowern, C., Colwell, R. K., Dunn, M., Dunn, R. R., … Yanega, G. (2013). Toward a mechanistic understanding of linguistic diversity. *BioScience*, *63*(7), 524–535. doi:10.1525/bio.2013.63.7.6

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). Cambridge University Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, 95–126.

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, 407–454.

Gimson, A. C. (1962). *An introduction to the pronunciation of English*. London: Edward Arnold.

Goldstein, H. (2011). *Multilevel statistical models*. Chichester: Wiley.

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society: Series A*, *170*(4), 941–954. doi:10.1111/j.1467-985X.2007.00491.x

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*(4), 455–467.

Gong, T., Shuai, L., & Zhang, M. (2014). Modelling language evolution: Examples and predictions. *Physics of Life Reviews*, *11*(2), 280–302. doi:10.1016/j.plrev.2013.11.009

Gooskens, C. (2007). The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development*, *28*(6), 445–467. doi:10.2167/jmmd511.0

Gooskens, C., Van Heuven, V. J., Van Bezooijen, R., & Pacilly, J. J. (2010). Is spoken Danish less intelligible than Swedish? *Speech Communication*, *52*(11), 1022–1037.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1304.

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, *426*(6965), 435–439. doi:10.1038/nature02029

Gray, R. D., & Jordan, F. M. (2000). Language trees support the express-train sequence of Austronesian expansion. *Nature*, *405*, 1052–1055.

Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language*, *32*(1), 109 – 115.

Greenberg, J. H. (1973). The typological method. In *Current trends in linguistics* (Vol. Diachronic, areal, and typological linguistics, pp. 149–193). The Hague: Mouton.

Haeserijn, W., Romijn, K., Geerts, G., de Rooij, J., & Van den Toorn, M., C. (1997). *Algemene nederlandse spraakkunst* (2nd ed.). Groningen / Deurne: Martinus Nijhoff / Wolters Plantyn. Retrieved from http://ans.ruhosting.nl/e-ans/index.html

Hahn, U. (2006). Similarity. In *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/0470018860.s00616/abstract

Hahn, U. (2014). Similarity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(3), 271–280. doi:10.1002/wcs.1282

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38. doi:10.1111/1467-9280.01415

Halle, M. (1973). Stress rules in english: A new version. *Linguistic Inquiry*, *4*(4), 451–464.

Hammarberg, B., Cenoz, J., Hufeisen, B., & Jessner, U. (2001). Roles of L1 and L2 in L3 production and acquisition. *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*, 21–41.

Hammarström, H., & O'Connor, L. (2013). Dependency-sensitive typological distance. In *Approaches to measuring linguistic differences*. Berlin: De Gruyter Mouton. Retrieved from http://www.degruyter.com/view/books/9783110305258/9783110305258.329/9783110305258.329.xml

Hancin-Bhatt, B. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research*, *10*(3), 241–269.

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, *29*(2), 163–180. doi:10.1177/0265532211421161

Haugen, E. I. (1966). Linguistics and language planning. *Sociolinguistics*, 50–71.

Haugen, E. I. (1969). *The Norwegian language in America: A study in bilingual behavior* (Vol. 1). Indiana University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hawkins, R. (2001). The theoretical significance of Universal Grammar in second language acquisition. *Second Language Research*, *17*(4), 345–367. doi:10.1177/026765830101700404

Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: the "failed functional features hypothesis." *Second Language Research*, *13*(3), 187–226. doi:10.1191/026765897671476153

Hayes, B. (2011). *Introductory phonology*. John Wiley & Sons.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. Taylor & Francis.

Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. University of Groningen. Retrieved from http://dissertations.ub.rug.nl/FILES/faculties/arts/2004/w.j.heeringa/titlecon.pdf

Heeringa, W., Kleiweg, P., Gooskens, C., & Nerbonne, J. (2006). Evaluation of string distance algorithms for dialectology. In *Proceedings of the Workshop on Linguistic Distances* (pp. 51–62). Retrieved from http://dl.acm.org/citation.cfm?id=1641984

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. doi:10.1017/S0140525X0999152X

Hill, J. H. (1978). Language contact systems and human adaptations. *Journal of Anthropological Research*, *34*(1), 1–26.

Hockett, C. F. (1958). *A course in modern linguistics*. New York: Macmillan.

Holman, E. W. (2010). Programs for calculating ASJP distance matrices (Version 2). Retrieved from http://email.eva.mpg.de/~wichmann/software.htm

Holman, E. W. (2011). The Newick reader: a program for producing a distance matrix from the branch lengths in a phylogenetic tree in Newick format (Version 1). Retrieved from http://email.eva.mpg.de/~wichmann/software.htm

Holman, E. W., Brown, C. H., Wichmann, S., Müller, A., Velupillai, V., Hammarström, H., … Egorov, D. (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, *52*(6), 841–875.

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., & Bakker, D. (2008). Explorations in automated language classification. *Folia Linguistica*, *42*(3-4), 331–354.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, amp048. doi:10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of l2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Psychology Press.

Huddleston, R., & Pullum, G. K. (2002). Preliminaries. In *The Cambridge Grammar of English.* (pp. 1–43). Cambridge: Cambridge University Press. Retrieved from http://catdir.loc.gov/catdir/samples/cam033/2001025630.pdf

Huettig, F. (in press). How literacy acquisition affects the illiterate mind. *Language & Linguistics Compass*.

Hufeisen, B., & Jessner, U. (2009). Learning and teaching multiple languages. In *Handbook of foreign language communication and learning* (pp. 109–137). Retrieved from http://books.google.nl/books?hl=en&lr=&id=3klgGNmjVpEC&oi=fnd&pg=PA109&dq=hufeisen+and+jessner,+2009&ots=zPa2szMj8h&sig=Wgyd4L0frZC0cIkeBEM5D-_bY1Q

Huff, P. (2010). Process_asjp: a script for generating subsets of ASJP data (Version .0.1). Retrieved from http://email.eva.mpg.de/~wichmann/software.htm

Hulstijn, J. H., Schoonen, R., de Jong, N. H., Steinel, M. P., & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of

Reference for Languages (CEFR). *Language Testing*, *29*(2), 203–221. doi:10.1177/0265532211419826

Hunley, K., Dunn, M., Lindström, E., Reesink, G., Terrill, A., Healy, M. E., … Friedlaender, J. S. (2008). Genetic and linguistic coevolution in Northern Island Melanesia. *PLoS Genetics*, *4*(10), e1000239. doi:10.1371/journal.pgen.1000239

Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, *60*(4), 877–925. doi:10.1111/j.1467-9922.2010.00577.x

Ionin, T., & Wexler, K. (2002). Why is "is" easier than "-s"?: acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research*, *18*(2), 95–136. doi:10.1191/0267658302sr195oa

Isphording, I. E., & Otten, S. (2011). *Linguistic distance and the language fluency of immigrants* (No. 274). Ruhr Economic Papers. Retrieved from http://www.econstor.eu/handle/10419/61444

Isphording, I. E., & Otten, S. (2013). The Costs of Babylon—Linguistic Distance in Applied Economics. *Review of International Economics*, *21*(2), 354–369. doi:10.1111/roie.12041

Isphording, I. E., & Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior & Organization*, *105*, 30–50. doi:10.1016/j.jebo.2014.03.027

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers). *The Journal of the Acoustical Society of America*, *126*(2), 866–877. doi:10.1121/1.3148196

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, *15*(2), 281–319.

Jaeger, T. F., Pontillo, D., & Graff, P. (2012). Comment on "Phonemic diversity supports a serial founder effect model of language expansion from Africa." *Science*, *335*(6072), 1042–1042. doi:10.1126/science.1215107

Jaeger, T. F., & Tily, H. (2011). On language "utility": processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 323–335. doi:10.1002/wcs.126

Jaensch, C. (2013). Third language acquisition: Where are we now? *Linguistic Approaches to Bilingualism*, *3*(1), 73–93. doi:10.1075/lab.3.1.04jae

Jager, G. (2003). Simulating language change with Functional OT. In S. Kirby (Ed.), *Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI* (pp. 52–61). Vienna. Retrieved from http://www.isrl.uiuc.edu/amag/langev/paper/jager03ESSLLI.html

Jakobson, R. (1941). *Kindersprache, Aphasie, und allgemeine Lautgesetze.* Uppsala: Almqvist and Wiksell.

Jakobson, R. (1968). *Child language: aphasia and phonological universals*. (A. Keller, Trans.). The Hague: Walter de Gruyter.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

Jessner, U. (2012). Dynamics of multilingualism. In *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd.

Jessner, U. (2014). On multilingual awareness or why the multilingual learner is a specific language learner. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism* (pp. 175–184). Springer.

Johnson. (2014, March 11). What is a foreign language worth? *The Economist*. Retrieved from http://www.economist.com/blogs/prospero/2014/03/language-study

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. doi:10.1016/0010-0285(89)90003-0

Joo, H.-R. (2003). Second language learnability and the acquisition of the argument structure of English locative verbs by Korean speakers. *Second Language Research*, *19*(4), 305–328.

Juola, P. (1998). Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics*, *5*(3), 206–213. doi:10.1080/09296179808590128

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*(02), 365–392.

Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition*, *2*(01), 37–57.

Kellerman, E. (1995). Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics*, *15*, 125–150. doi:10.1017/S0267190500002658

Kellerman, E., & Sharwood Smith, M. (1986). *Crosslinguistic influence in second language acquisition*. New York: Pearson College Division.

Kessler, B. (2005). Phonetic comparison algorithms. *Transactions of the Philological Society*, *103*(2), 243–260.

Kim, M.-H., & Lee, H.-H. (2010). Linguistic and nonlinguistic factors determining proficiency of English as a foreign language: a cross-country analysis. *Applied Economics*, *42*(18), 2347–2364. doi:10.1080/00036840701857960

Klein, E. (1995). Second versus third language acquisition: Is there a difference? *Language Learning*, *45*(3), 419–466. doi:10.1111/j.1467-1770.1995.tb00448.x

Klein, W., & Perdue, C. (1997). The basic variety (or: couldn't natural languages be much simpler?). *Second Language Research*, *13*(4), 301–347. doi:10.1191/026765897666879396

Koda, K. (1989). The effects of transferred vocabulary knowledge on the development of l2 reading proficiency. *Foreign Language Annals*, *22*(6), 529–540. doi:10.1111/j.1944-9720.1989.tb02780.x

Kominski, R. (1989). How good is "how well"? An examination of the census english-speaking ability question. In *American Statistical Association 1989 Proceedings of the social statistics section* (pp. 333–338). Retrieved from http://www.census.gov/population/socdemo/language/ASApaper1989.pdf

Kondrak, G. (2000). A new algorithm for the alignment of phonetic sequences. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* (pp. 288–295). Seattle. Retrieved from http://dl.acm.org/citation.cfm?id=974343

Kondrak, G. (2001). Identifying cognates by phonetic and semantic similarity. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 103–110). Pittsburgh, PA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1073350

Kondrak, G., & Sherif, T. (2006). Evaluation of several phonetic similarity algorithms on the task of cognate identification. In *Proceedings of the Workshop on Linguistic Distances* (pp. 43–50). Sydney: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1641983

Kortmann, B., & Szmrecsanyi, B. (2012). *Linguistic complexity: Second language acquisition, indigenization, contact*. Berlin: Mouton de Gruyter.

Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*(5940), 611–612. doi:10.1126/science.1173947

Kramsch, C. (2012). Imposture: A late modern notion in poststructuralist SLA research. *Applied Linguistics*, ams051.

Kreft, I. G. G., & Leeuw, J. D. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Kubota, R. (1998). An investigation of L1–L2 transfer in writing among Japanese university students: Implications for contrastive rhetoric. *Journal of Second Language Writing*, *7*(1), 69–100. doi:10.1016/S1060-3743(98)90006-6

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. doi:10.3758/BF03212211

Kuijper, H., Bergsma, A., & Bechger, T. (2004). *Staatsexamen NT2. Het gewenste niveau. Deel 1 behoeftepeiling programma II [State Examination for Dutch as a second language. The desired level. Part I a needs inquiry for program II]*. Arnhem, The Netherlands: Central Institute for Test Development.

Kusters, W. (2003). *Linguistic complexity: The influence of social change on verbal inflection*. Universiteit Leiden. Retrieved from

http://www.lotpublications.nl/publish/articles/000501/bookpart.pdf
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). (Version 2.0-11). Retrieved from http://CRAN.R-project.org/package=lmerTest

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (2007). Transmission and diffusion. *Language*, *83*, 344–387.

Ladd, D. R., Roberts, S. G., & Dediu, D. (2015). Correlational studies in typological and historical linguistics. *Annual Review of Linguistics*, *1*(1), null. doi:10.1146/annurev-linguist-030514-124819

Ladefoged, P., & Maddieson, I. (1998). *The sounds of the world's languages*. Oxford: Blackwell.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Michigan: University of Michigan Press.

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, *14*(4), 359–375. doi:10.1191/026765898672500216

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A*, *172*(3), 537–554. doi:10.1111/j.1467-985X.2008.00577.x

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. doi:10.3758/s13428-011-0146-0

Lenneberg, E. H. (1967). *Biological foundations of language* (Vol. 68). Wiley New York. Retrieved from http://tocs.ulb.tu-darmstadt.de/123135893.pdf

Levinson, S. C. (1979). Activity types and language. *Linguistics*, *17*(5-6), 365–400. doi:10.1515/ling.1979.17.5-6.365

Levinson, S. C. (1980). Speech act theory: The state of the art. *Language Teaching*, *13*(1-2), 5–24. doi:10.1017/S0261444800008661

Levinson, S. C. (2000). Yélî dnye and the theory of basic color terms. *Journal of Linguistic Anthropology*, *10*(1), 3–55.

Levinson, S. C., & Gray, R. D. (2012). Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences*, *16*(3), 167–173. doi:10.1016/j.tics.2012.01.007

Lev, M., Stark, T., & Chang, W. (2012). *South American phonological inventory database*. University of California. Retrieved from http://linguistics.berkeley.edu/ saphon/en/

Levy, R., & Daumé, H. (2011). Computational methods are invaluable for typology, but the models must match the questions. *Linguistic Typology*, *15*(2), 393–399.

Lewis, P., Simons, G., & Fennig, C. (Eds.). (2013). *Ethnologue: Languages of the world* (17th ed.). Dallas, Texas: SIL International. Retrieved from http://www.ethnologue.com

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. doi:10.1016/j.tics.2006.08.007

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. doi:10.1016/j.jml.2010.07.003

Lightfoot, D. (1999). *The development of language: Acquisition, change and evolution*. Blackwell: Oxford. Retrieved from http://www.isrl.uiuc.edu/ amag/langev/paper/lightfoot99book.html

Lindqvist, C. (2010). Inter- and intralingual lexical influences in advanced learners' French L3 oral production. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(2-3), 131–157. doi:10.1515/iral.2010.007

Lopez, M. H. (1999). Does speaking a second language affect labor market outcomes? Evidence from national adult literacy survey of 1992. *Unpublished Paper, School of Public Affairs, University of Maryland*.

Lüdeling, A., Walter, M., Kroymann, E., & Adolphs, P. (2005). Multi-level error annotation in learner corpora. *Proceedings of Corpus Linguistics 2005*, 15–17.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, *5*(1), e8559. doi:10.1371/journal.pone.0008559

Lupyan, G., & Dale, R. (2015). The role of adaptation in understanding linguistic diversity. In R. De Busser & R. J. LaPolla (Eds.), *Language structure and environment: Social, cultural, and natural factors.* (pp. 287–316). Amsterdam: John Benjamins Publishing Company.

Luyckx, K., Kloots, H., Coussé, E., & Gillis, S. (2007). Klankfrequenties in het Nederlands. In *Tussen taal, spelling en onderwijs : essays bij het emeritaat van Frans Daems* (pp. 141–154). Academia Press. Retrieved from http://hdl.handle.net/1854/LU-430290

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *The Modern Language Journal*, *90*(3), 320–337.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, I. (2011a). Consonant inventories. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Digital Library. Retrieved from http://wals.info/chapter/1

Maddieson, I. (2011b). Vowel quality inventories. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Digital Library. Retrieved from http://wals.info/chapter/2

Major, R. C. (2008). Transfer in second language phonology. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (Vol. 36, pp. 63–94). Amsterdam: John Benjamins Publishing.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229–258.

McArthur, E. K., & Siegel, P. M. (1983). Developments in the measurement of English language proficiency. In *Proceedings of the American Statistical Association (Social Statistics Section)* (pp. 373–378).

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*(03), 395–423.
McMahon, A. M. S., & McMahon, R. (2005). *Language classification by numbers*. Oxford University Press.
McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126. doi:10.1207/s15516709cog0000_79
McWhorter, J. H. (2002). What happened to English? *Diachronica*, *19*(2), 217–272. doi:10.1075/dia.19.2.02wha
McWhorter, J. H. (2007). *Language interrupted: Signs of non-native acquisition in standard language grammars*. New York: Oxford University Press.
McWhorter, J. H. (2011). *Linguistic simplicity and complexity: Why do languages undress?*. Boston: De Gruyter Mouton.
Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, *41*(4), 473–497.
Michaels, D. (1974). Sound replacements and phonological systems. *Linguistics*, *12*(126), 69–82. doi:10.1515/ling.1974.12.126.69
Miestamo, M., Sinnemäki, K., & Karlsson, F. (2008). *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins Publishing.
Mira, J., & Paredes, A. (2005). Interlinguistic similarity and language death dynamics. *Europhysics Letters*, *69*(6), 1031–1034. doi:10.1209/epl/i2004-10438-4
Mitchell, R., & Myles, F. (2004). *Second language learning theories*. London: Arnold Publishers.
Moran, S., & Blasi, D. (2014). Cross-linguistic comparison of complexity measures in phonological systems. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity*. oxford: Oxford University Press.
Moran, S., McCloy, D., & Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language*, *88*(4), 877–893.
Moran, S., McCloy, D., & Wright, R. (Eds.). (2014). *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://phoible.org/

Moran, S., & Wright, R. (2009). *Phonetics information base and lexicon (PHOIBLE)*. Seattle, WA. Retrieved from http://phoible.org/

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge: Cambridge University Press.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 193–219). Amsterdam: John Benjamins.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. doi:10.1111/j.1467-1770.1995.tb00963.x

Murphy, S. (2005). Second language transfer during third language acquisition. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, *3*(1). Retrieved from http://journal.tc-library.org/index.php/tesol/article/view/23

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. doi:10.1111/j.2041-210x.2012.00261.x

Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1829–1836. doi:10.1098/rstb.2011.0216

Nicholls, D. (2003). The cambridge learner corpus: Error coding and analysis for lexicography and elt. In *Proceedings of the Corpus Linguistics Conference* (pp. 572–581). Retrieved from http://ucrel.lancs.ac.uk/publications/cl2003/papers/nicholls.pdf

Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, *417*(6889), 611–617. doi:10.1038/nature00771

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.

Odlin, T. (2005). Crosslinguistic influence and conceptual transfer: what are the concepts? *Annual Review of Applied Linguistics*, *25*, 3–25. doi:10.1017/S0267190505000012

O'Grady, R. T., Goddard, I., Bateman, R. M., DiMichele, W. A., Funk, V. A., Kress, W. J., … Cannell, P. F. (1989). Genes and tongues. *Science*, *243*, 1651.

Oldenkamp, L. (2013). *The trouble with inflection of adult learners of dutch: A study on the L1-L2 interplay of morphosyntactic and phonetic-phonological factors* (PhD Thesis). Radboud University Nijmegen.

O'Rourke, B., & Pujolar, J. (2013). From native speakers to new speakers-problematizing nativeness in language revitalization contexts. *Histoire Épistémologie Langage*, *35*(2), 47–67.

Ostler, N. (2005). *Empires of the word: a language history of the world*. New York: Harper Collins.

Oudeyer, P.-Y. (2001). Origins and learnability of syllable systems: A cultural evolutionary model. In C. P, F. C, J. K. Hao, E. Lutton, & S. M (Eds.), *Artificial Evolution, LNCS 2310* (pp. 143–155). Springer Verlag. Retrieved from http://www.isrl.uiuc.edu/amag/langev/paper/oudeyerea2001.html

Oudeyer, P.-Y. (2005). How phonological structures can be culturally selected for learnability. *Adaptive Behavior*, *13*(4), 269–280. doi:10.1177/105971230501300407

Pajak, B. (2012). *Inductive inference in non-native speech processing and learning*. University of California at San Diego. Retrieved from http://www.bcs.rochester.edu/people/bpajak/pdfs/Pajak_2012_diss.pdf

Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2014). *Learning additional languages as hierarchical probabilistic inference: insights from L1 processing*. Rochester. Retrieved from https://www.academia.edu/6272547/Pajak_B._Fine_A.B._Kleinschmidt_D._and_Jaeger_T._F._resubmitted._Learning_additional_languages_as_hierarchical_probabilistic_inference_insights_from_L1_processing

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins Publishing.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306–338. doi:10.1016/j.cognition.2010.11.001

Pica, T. (1983). Adult acquisition of english as a second language under different conditions of exposure. *Language Learning*, *33*(4), 465–497. doi:10.1111/j.1467-1770.1983.tb00945.x

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2013). *nlme: Linear and nonlinear mixed effects models*.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press. Retrieved from http://www.isrl.uiuc.edu/amag/langev/paper/pinker84languageLearnability.html

Piske, T., Flege, J. E., MacKay, I. R., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, *59*(1), 49–71.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*(2), 191–215. doi:10.1006/jpho.2001.0134

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*(2), 103–133. doi:10.1191/026765800677556046

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275*(5306), 1604–1610.

Rafferty, A. N., Griffiths, T. L., & Ettlinger, M. (2011). Exploring the relationship between learnability and linguistic universals. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 49–57). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2021096.2021102

Ragin, C. (2004). Turning the tables: how case-oriented research challenges variable oriented research. In H. Brady & D. Collier (Eds.), *Rethinking social inquiry: Diverse tools, shared standards* (pp. 123–38). Lanham, MD: Rowman & Littlefield.

Rasbash, J., & Browne, W. J. (2008). Non-hierarchical multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 301–334). New York, NY: Springer.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*, *18*(4), 321–349. doi:10.2307/1165158

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

R Core Team. (2013). *R: A language and environment for statistical computing*. Retrieved from http://www.r-project.org/

Rehm, G., & Uszkoreit, H. (2013). Multilingual Europe: Facts, challenges, opportunities. In G. Rehm & H. Uszkoreit (Eds.), *META-NET strategic research agenda for multilingual Europe 2020* (pp. 12–18). Heidelberg: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-36349-8_2

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon, UK: Multilingual Matters.

Ritchie, W. C. (1968). On the explanation of phonic interference. *Language Learning*, *18*(3-4), 183–197.

Rivers, W. P., & Golonka, E. M. (2009). Third language acquisition theory and practice. In M. H. Longessor & C. J. Directoressor (Eds.), *The handbook of language teaching* (pp. 250–266). Wiley-Blackwell.

Roberts, S., & Winters, J. (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PLoS ONE*, *8*(8), e70902. doi:10.1371/journal.pone.0070902

Robinson, G. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, *6*(1), 15–32.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*(01), 27–67. doi:10.1017/S0272263100014674

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, *27*(02), 235–268.

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(2-3), 245–273.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*(1), 107–127. doi:10.1177/0267658310386439

Ruhlen, M. (1991). *A Guide to the World's Languages: Classification*. Stanford University Press.

Saiz, A., & Zoido, E. (2002). The returns to speaking a second language. In *Working paper NO. 02-16*. Philadelphia: Fed's Research Department.

Sampson, G., Gil, D., & Trudgill, P. (2009). *Language complexity as an evolving variable*. Oxford: Oxford University Press.

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, *15*(1), 157–166. doi:10.1017/S1366728910000623

Schepens, J., Van der Slik, F., & Van Hout, R. (submitted). The L2 impact on acquiring Dutch as a L3: the L2 distance effect. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed Effects Regression Models in Linguistics*. Berlin: Springer.

Schepens, J., Van der Slik, F., & Van Hout, R. (2013a). Learning complex features: A morphological account of l2 learnability. *Language Dynamics and Change*, *3*(2), 218–244. doi:10.1163/22105832-13030203

Schepens, J., Van der Slik, F., & Van Hout, R. (2013b). The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In L. Borin & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 199–230). Berlin: De Gruyter Mouton.

Schmidt, R. W. (1977). *Sociolinguistic variation and language transfer in phonology*. Ontario Institute for Studies in Education. Retrieved from http://nflrc.hawaii.edu/PDFs/SCHMIDT%20Sociolinguistic%20variation%20and%20language%20transfer.pdf

Schumann, J. H., Crowell, S. E., Jones, N. E., Lee, N., & Schuchert, S. A. (2004). *The neurobiology of learning: Perspectives from second language acquisition*. Routledge.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*(1), 40–72.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*(1-4), 209–232.

Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(03), 243–261.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shi, Y., Leite, W., & Algina, J. (2010). The impact of omitting the interaction between crossed factors in cross-classified random effects modelling. *British Journal of Mathematical and Statistical Psychology*, *63*(1), 1–15. doi:10.1348/000711008X398968

Siegel, P., Martin, E., & Bruno, R. (2001). Language use and linguistic isolation: Historical data and methodological issues. *US Census Bureau: February*, *12*. Retrieved from https://www.census.gov.edgekey.net/hhes/socdemo/language/data/census/li-final.pdf

Singleton, D. (1987). Mother and other tongue influence on learner French. *Studies in Second Language Acquisition*, *9*(03), 327–345. doi:10.1017/S0272263100006719

Slabakova, R. (2006). Learnability in the second language acquisition of semantics: A bidirectional study of a semantic parameter. *Second Language Research*, *22*(4), 498–523.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of Child Language Development* (Vol. 1, pp. 175–208). New York: Holt, Rinehart & Winston. Retrieved from http://ihd.berkeley.edu/Slobin-Language%20Acquisition/(1973)%20Slobin%20-%20Cognitive%20prerequisites%20for%20acquisition%20of%20grammar.pdf

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444–9.

Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE.

Snow, M. S. (1998). Economic, statistical, and linguistic factors affecting success on the test of English as a foreign language (TOEFL). *Information Economics and Policy*, *10*(2), 159–172. doi:10.1016/S0167-6245(97)00018-8

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2003). Unsupervised context sensitive language acquisition from a large corpus. In *NIPS-2003*. Retrieved from http://www.isrl.uiuc.edu/amag/langev/paper/solan03nips.html

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(33), 11629–11634. doi:10.1073/pnas.0409746102

Spiess, A. N. (2013). qpcR: Modelling and analysis of real-time PCR data (Version 1.3-7.1). Retrieved from http://cran.r-project.org/web/packages/qpcR/index.html

Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, *8*(4), 339–356.

Stevens, G. (1986). Sex differences in language shift in the United States. *Sociology and Social Research*, *71*(1), 31–36.

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, *115*(4), 1791–1807. doi:10.1121/1.1687832

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. *Phonology and Second Language Acquisition*, 153–191.

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society*, 452–463.

Szmrecsanyi, B. (2004). On operationalizing syntactic complexity. *Jadt-04*, *2*, 1032–1039.

Szmrecsanyi, B., & Kortmann, B. (2009). Between simplification and complexification: non-standard varieties of English around the world. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language* *complexity as an evolving variable* (pp. 64–79). Oxford: Oxford University Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(04), 629–640. doi:10.1017/S0140525X01000061

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. doi:10.1126/science.1192788

Ternes, E. (1976). *Probleme der kontrastiven Phonetik*. Buske Verlag.

Tesar, B., & Smolensky, P. (1996). Learnability in optimality theory (long version). *Report No. JHU_CogSci_96_3. Baltimore, MD: Johns Hopkins University*. Retrieved from http://mind.cog.jhu.edu/faculty/smolensky.html/Site/docs/TesarSmolensky96TRLearnabilityLong.pdf

Tesar, B., & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*, *29*(2), 229–268.

Thomas, J. (1988). The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual & Multicultural Development*, *9*(3), 235–246.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.

Tily, H., Frank, M. C., & Jaeger, T. F. (2011). The learnability of constructed languages reflects typological patterns. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1364–1369).

Towell, R., & Hawkins, R. D. (1994). *Approaches to second language acquisition*. Clevedon, UK: Multilingual Matters.

Trask, R. L. (2000). *Time depth in historical linguistics*. The McDonald Institute for Archaeological Research.

Trudgill, P. (1974). Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. *Language in Society*, *3*(02), 215–246. doi:10.1017/S0047404500004358

Trudgill, P. (1983). *On dialect: Social and geographical perspectives*. Oxford: Basil Blackwell.

Trudgill, P. (2001). Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology*, *5*(2-3), 371–374.

Trudgill, P. (2002). *Sociolinguistic variation and change*. Edinburgh: Edinburgh University Press.
Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
Tsukada, K., Hirata, Y., & Roengpitya, R. (2014). Cross-language perception of japanese vowel length contrasts: comparison of listeners from different first language backgrounds. *Journal of Speech Language and Hearing Research*, 805. doi:10.1044/2014_JSLHR-S-12-0416
Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. doi:10.1037/0033-295X.84.4.327
Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. *Mind and Context in Adult Second Language Acquisition*, 141–178.
UNESCO. (2011). *The World Bank, Institute for Statistics, School enrollment, secondary (% gross)*. Retrieved January 3, 2013, from http://data.worldbank.org/indicator/SE.SEC.ENRR
Van der Slik, F. (2010). Acquisition of Dutch as a second language. *Studies in Second Language Acquisition*, *32*(03), 401–432. doi:10.1017/S0272263110000021
Van der Slik, F., Van Hout, R., & Schepens, J. (submitted). The gender gap in second language acquisition: Robust gender differences in the acquisition of Dutch as a second language among immigrants from 90 countries speaking 49 first languages.
Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, *62*, 148–171. doi:10.1111/j.1467-9922.2012.00710.x
Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PLoS ONE*, *8*(7), e69172. doi:10.1371/journal.pone.0069172
Van Tubergen, F., & Kalmijn, M. (2005). Destination language proficiency in cross national perspective: A study of immigrant groups in nine western countries. *American Journal of Sociology*, *110*(5), 1412–1457. doi:10.1086/428931
Van Tubergen, F., & Kalmijn, M. (2009). Language proficiency and usage among immigrants in the Netherlands: Incentives or opportunities? *European Sociological Review*, *25*(2), 169–182. doi:10.1093/esr/jcn043

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Journal of Second Language Writing*, *18*(4), 235–250. doi:10.1016/j.jslw.2009.06.003

Verhelst, N. D., & Glas, C. A. W. (1995). The one parameter logistic model: OPLM. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 215–238). New York: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4612-4230-7_12

Von Humboldt, W. (1836). *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Dümmler.

Weinreich, U. (1963). *Languages in Contact*. The Hague: Mouton.

Wichmann, S. (In press). Diachronic stability and typology. In C. Bowern & B. Evans (Eds.), *Handbook of Historical Linguistics*. London: Routledge.

Wichmann, S., & Holman, E. W. (2009). Population size and rates of language change. *Human Biology*, *81*(2-3), 259–274. doi:10.3378/027.081.0308

Wichmann, S., Holman, E. W., Bakker, D., & Brown, C. H. (2010). Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and Its Applications*, *389*(17), 3632–3639. doi:10.1016/j.physa.2010.05.011

Wichmann, S., Müller, A., Velupillai, V., Brown, C. H., Holman, E. W., Brown, P., … others. (2010). The ASJP database (version 13). Retrieved from http://email.eva.mpg.de/wichmann/languages.htm

Wichmann, S., Stauffer, D., Schulze, C., & Holman, E. W. (2008). Do language change rates depend on population size? *Advances in Complex Systems*, *11*(03), 357–369. doi:10.1142/S0219525908001684

Wieling, M., Bloem, J., Mignella, K., Timmermeister, M., & Nerbonne, J. (2014). Measuring foreign accent strength in English: Validating Levenshtein distance as a measure. *Language Dynamics and Change*, *4*(2), 253–269.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), e23613. doi:10.1371/journal.pone.0023613

Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics*, *19*(3), 295–333.

Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, *30*(2), 231–252. doi:10.1177/0265532212456968

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, *27*(2), 147–170. doi:10.1177/0265532209349465

Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011). A new dataset and method for automatically grading ESOL texts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (pp. 180–189). Retrieved from http://www.aclweb.org/anthology/P/P11/P11-1019.pdf

Samenvatting in het Nederlands

Veel volwassen mensen leren een tweede of derde taal (T2, T3, etc.) nadat ze één of meerdere talen eerder hebben geleerd (T1, T2, etc.), inclusief uiteraard hun moedertaal (T1). De nieuwe talen zullen moeilijker of gemakkelijker te leren zijn dan de eerdere talen, misschien zelfs mooier of belangrijker zijn, maar in elk geval zijn ze anders dan de eerder geleerde talen. De diversiteit aan talen is groot wereldwijd en dat lijkt te zorgen voor verschillen in de leerbaarheid van additionele talen bij volwassenen, afhankelijk van hun taalachtergrond. Leerbaarheid van een additionele taal kan omschreven worden als de moeite die het kost om die taal te leren, gegeven de eerder geleerde talen.

Dit proefschrift onderzoekt de effecten van verschillen in taalachtergrond (zowel de moedertaal, T1, als wellicht de tweede taal, T2) op de leerbaarheid van het Nederlands (als additionele taal) bij volwassenen. De kernhypothese is dat taalafstand, oftewel de mate waarin talen van elkaar verschillen, de leerbaarheid van additionele talen beïnvloedt. Hoe kleiner de afstand, hoe minder moeite het kost en dus hoe beter de leerbaarheid. Maar hoe kunnen we de taalafstand nu bepalen? Wat voor taalafstandsmaten zijn bruikbaar? Bijvoorbeeld, als T1 sprekers van het Engels het Nederlandse woord ‘diversiteit’ leren, zal de leerbaarheid relatief hoog zijn omdat betekenis en vorm overeen lijken te komen met het Engelse woord ‘diversity’. Voor het woord ‘mooi’ daarentegen, zal de leerbaarheid lager uitvallen, want bij ‘beautiful’ is het Engels gaan leunen bij het Frans. In dit proefschrift worden uiteenlopende afstandsmaten ontwikkeld en gebruikt. De verschillende afstandsmaten laten zien wat voor verschillen tot problemen leiden en wat voor overeenkomsten nuttig zijn voor leerders. De algemene conclusie luidt dat des te groter de structurele verschillen tussen de verschillende betrokken achtergrondtalen (T1 plus eventuele T2s) en het Nederlands als doeltaal zijn, des te groter de leerbaarheidsproblemen. Dat geldt voor elk van de drie ontwikkelde maten: de lexicale, de morfologische en de fonologische afstandsmaat.

Leerbaarheid bij volwassenen

Door de jaren heen hebben taalwetenschappers steeds weer beargumenteerd waarom leerders kunnen profiteren van taalcursussen die zijn afgestemd op taalachtergrond (Lado, 1957; Cook, 2013). Toch weten we nog onvoldoende over de manier waarop leerders met taalverschillen omgaan. In de jaren zestig is men vanuit de Contrastieve Taalkunde begonnen met het systematisch vergelijken van talen (Weinreich, 1963) en met onderzoek naar de relevantie van taalverschillen voor tweede-taalverwerving. Daarna is echter tot vrij recent de aandacht terecht gekomen op de universele leerproblemen van leerders van een tweede taal. Tegenwoordig staat de rol van taalachtergrond en de taalverschillen weer meer in de belangstelling, bijvoorbeeld door onderzoek naar de invloed van de T1 op de verwerving van T2-klanken (Flege, 2005) of T2-grammatica (Ionin & Montrul, 2010). Dit soort van onderzoek vergelijkt de verwerving van een specifieke structuur in een doeltaal tussen groepen van leerders met verschillende T1s. Het is vaak echter niet mogelijk om veel groepen van leerders met elkaar te vergelijken. Toch laat een aantal studies zien dat het bestuderen van meer verschillende leerders tot nieuwe inzichten kan leiden met betrekking tot de rol van leeftijd (Hakuta, Bialystok, and Wiley, 2003). Er zijn geen grootschalige studies uitgevoerd naar de manier waarop uiteenlopende groepen leerders met taalverschil omgaan en wat voor verschillen belangrijk zijn.

Al het onderzoek in dit proefschrift maakt gebruik van een dataset die mij in staat stelt om verschillen tussen een groot aantal groepen leerders te bestuderen. Ik gebruik als maat voor leerbaarheid de spreekvaardigheidsbeoordelingen van het staatsexamen Nederlands als Tweede Taal. Dit zijn testresultaten die bestaan uit de beoordelingsscores voor meer dan 50.000 deelnemers aan dit examen tussen 1995 en 2010. Er zijn genoeg deelnemers om 74 verschillende T1s en 35 verschillende T2s te vergelijken. De deelnemers moeten deze toets meestal maken om aan een Nederlandse universiteit of hogeschool te kunnen studeren of voor een Nederlands bedrijf te mogen werken. De spreekvaardigheid van de deelnemers wordt beoordeeld door speciaal opgeleide beoordelaars op zowel vorm als inhoud. Er ontstaat een score die zowel woordgebruik, grammatica, en uitspraak meeneemt. Ik duid de dataset aan met STEX. STEX is een unieke dataset omdat het hiermee mogelijk wordt om veel verschillende factoren met uiteenlopende waardes in het onderzoek te betrekken. De deelnemers hebben bijvoorbeeld een breed scala aan leeftijden ten tijde van het examen en ten tijde van immigratie naar Nederland. Hierdoor kan ik dus kijken naar effecten van zowel leeftijd als ook blootstelling aan het Nederlands, oftewel de tijd tussen immigratie en examen. De meeste deelnemers in STEX hebben namelijk vrijwillig een vragenlijst ingevuld ten tijde van het examen. Naast de vraag naar de immigratiedatum, vraagt de lijst onder andere ook naar land van herkomst, T1, beste T2, jaren dagonderwijs, en uren Nederlandse les. De antwoorden op deze vragenlijst stellen me in staat om onderzoek naar de effecten van taalachtergrond te controleren voor een reeks van andere factoren die ook van invloed (kunnen) zijn.

Taalafstand

Meertaligheidsonderzoek heeft lange tijd niet de beschikking gehad over metingen van taalverschil tussen vele verschillende talen. Recentelijk is het echter mogelijk geworden om deze metingen toch te verrichten, vooral op grond van recentelijk taaltypologisch onderzoek waarbij databanken met gegevens over grote aantallen talen beschikbaar zijn gekomen. Ik ontwikkel op grond van die databankenmaten van morfologische en fonologische taalafstand en daarnaast gebruik ik eerder ontwikkelde lexicale taalafstandsmaten. Voor de nieuwe afstandsmaten heb ik informatie gebruikt uit de taaltypologie, waaronder de ´Wereldatlas van Taalstructuren´ (World Atlas of Language Structures, WALS; Dryer & Haspelmath, 2011) en de ´Fonetische Informatie Database´ (Phonetic Information Base and Lexicon, PHOIBLE; Moran, 2014). Deze data geven inzicht in de structurele overeenkomsten en verschillen tussen talen.

De literatuur bevat verschillende ideeën over wat taalverschillen zijn en wat voor afstandsmaten deze verschillen kunnen meten. Zo zijn er ideeën voortgekomen vanuit onderzoek naar het lexicon, de morfologie, en de fonologie. In dit proefschrift betrek ik in mijn afstandsmaten de rol van evolutionaire verandering van woorden, de typologie van bepaalde morfologische constructies, en de abstracte articulatorische (fonologische) eigenschappen van klanken. In al deze domeinen is het mogelijk om afstanden te bepalen. Een woord, regel, of klank in een andere taal kan verschillend, gelijk, en alles daar tussen in zijn. Bij verschillen moet er vaak iets worden bijgeleerd, wat een verhoging in complexiteit is met betrekking tot eerder geleerde structuur. Maar soms moet er ook iets worden afgeleerd, bijvoorbeeld door een bepaalde klank of woord niet meer te gebruiken of door van een drievoudig geslachtssysteem (zoals het Duitse "der", "die", "das") naar een tweevoudig geslachtssysteem te gaan (het Nederlandse "de" en "het").

Het proefschrift omvat vijf onderzoeken naar de relatie tussen taalafstand en leerbaarheid. Hoofdstuk 2, 3, en 4 behandelen studies naar de invloed van lexicale, morfologische, en fonologische afstand. Hoofdstuk 5 en 6 laten zien dat deze afstanden ook van invloed zijn voor de T2 naast de T1 voor het leren van Nederlands als additionele taal (T3).

Resultaten

Effecten van T1 Taalafstand

Lexicale Afstand. De eerste studie (Hoofdstuk 2) presenteert de bevindingen van twee recente studies naar lexicale afstand naar de studie van T2 leerbaarheid. Tot op heden maakten vergelijkbare studies van T2 leerbaarheid nog geen gebruik van deze afstandsmaten. Ik onderzoek eerst welke van de twee lexicale afstandsmaten de variatie in spreekvaardigheidsscores beter kan verklaren. Hiervoor gebruik ik multi-level modellen. Deze modellen ontleden variatie in spreekvaardigheidsscores in zowel vaste als steekproef-afhankelijke variatiebronnen in hetzelfde lineaire regressiemodel. Ik verdeel de algehele variatie op tussen landen, talen, en individuele variatie.

Vervolgens stel ik vast dat de maat die relatief grote verschillen tussen Germaanse talen voorspelt ten opzichte van andere Indo-Europese talen meer variatie in spreekvaardigheidsscores kan voorspellen. Beiden maten bepalen op een andere manier de hoeveelheid van evolutionaire verandering tussen talen. Kortweg, de lexicale maat die terugkerende klankovereenkomsten gebruikt werkt beter dan de maat die het aantal klankovereenkomsten zelf gebruikt.

Morfologische Afstand. De tweede studie (Hoofdstuk 3) onderzoekt het effect van morfologische taalverschillen op T2 leerbaarheid en laat zien dat een verhoogde T2 morfologische complexiteit tot meer leerproblemen leidt. Dit effect wordt ook gebruikt door taaltypologen in verklaringen voor complexiteitsverschillen tussen talen (Dahl, 2004; Trudgill, 2011). De rol van de mate van taalverschil in relatie tot dit effect is nog niet eerder bestudeerd. Een tweede innovatie is dat deze studie een morfologische taalafstandsmaat oplevert waarmee taalafstanden tussen talen die niet direct aan elkaar verwant zijn bepaald kunnen worden.

Fonologische Afstand. De derde studie (Hoofdstuk 4) onderzoekt het effect van klankverschillen op T2 leerbaarheid. Volwassen leerders van een T2 blijven meestal een accent houden. Dit fenomeen kenmerkt het probleem dat nieuwe klanken veroorzaken. Het aantal nieuwe klanken en hoe deze nieuwe klanken verschillen hangt af van de klanken in eerder geleerde talen. Niet alle talen maken gebruik van dezelfde klanken. De selectie aan klanken waarmee een taal haar lexicon opbouwt is onderhevig aan taalveranderingsprocessen. Hierdoor ontstaat diversiteit in klanksystemen. Zo staat bijvoorbeeld het Nederlandse klinkersysteem als relatief ingewikkeld bekend. Wat maakt het moeilijk om een T2 klanksysteem te leren? Mijn vermoeden was dat het produceren van T2 klanken lastig gemaakt wordt doordat de articulatorische patronen waarmee T1 klanken geproduceerd kunnen worden al ingeslepen zijn. Om dit te testen heb ik gemeten hoeveel nieuwe articulatorische patronen leerders moeten toevoegen aan hun reeds verworven patronen om alle Nederlandse klanken te kunnen produceren. Het bleek dat het aantal nieuwe klanken met T2 leerbaarheid correleert, maar dat het aantal nieuwe patronen dat nodig is voor deze klanken nog belangrijker is.

Effecten van T2 Taalafstand

In Hoofdstukken 2, 3, 4 hebben we effecten van het al kunnen spreken van een andere tweede taal naast het Nederlands niet nader proberen te verklaren. In Hoofdstukken 5 en 6 ligt hier de focus wel. Ongeveer 80% van de deelnemers in STEX spreken een tweede taal (T2) en hebben vervolgens aangegeven welke taal ze naast hun moedertaal (T1) en het Nederlands (T3) als beste beheersen.

T1 bij T2 Interactie Effecten. De vierde studie (Hoofdstuk 5) onderzoekt hoe keuzes voor bepaalde statistische procedures de interpretatie van de rol van de T2 beïnvloeden. Een belangrijke assumptie van multi-level modellen is de aanname dat variatie binnen de ene steekproef, bijvoorbeeld voor de T1s, niet afhangt van variatie binnen de andere steekproef, bijvoorbeeld voor de T2s. Dit hoofdstuk toont aan dat de combinaties van T1 en T2 een effect hebben op de T3 leerbaarheid, maar vooral ook dat de T1 en T2 afzonderlijk van sterkere invloed zijn.

Effecten van T2 Taalafstand. Binnen de T3 literatuur bestaan er verschillende opvattingen over de manier waarop leerders gebruik maken van hun T1 en T2 tijdens het leren van een T3. Zo bestaan er theorieën die voorspellen dat de belangrijkste taal 1) de taal met de kleinste afstand is, 2) de taal die als tweede taal geleerd is, 3) de eerste taal kan zijn maar dat de tweede taal ook belangrijk is. Daarnaast bestaat er de opvatting dat het meertalig zijn een inherent positief effect heeft op het leren van nog meer talen. Het onderzoek in dit hoofdstuk levert bewijs voor de autonome invloed van T2 taalafstand, naast een sterker en autonoom effect van de T1. Zowel de lexicale als morfologische taalafstanden verklaren variatie tussen zowel T1s als ook T2s, onafhankelijk van het eventueel weglaten van de eentaligen. De effecten voor taalafstand zijn additief. Combinatie-specifieke variatie kan niet worden verklaard door systematische interacties tussen T1- en T2-taalafstanden. Meertalig zijn is over het algemeen voordelig, hoewel een ver verwijderde T2 niet altijd voordeliger is dan eentalig zijn. Deze bevindingen zijn consistent met theorie nummer 3. Mijn onderzoek voegt daar aan toe dat taalafstand belangrijk is.

Discussie en Conclusies

Samen geven de behandelde onderzoeken een beeld van de effecten van taalafstand op de T2-leerbaarheid van het Nederlands in het lexicale, morfologische, en fonologische domein (Hoofdstukken 2, 3, en 4), op T3-leerbaarheid (Hoofdstukken 5 en 6), en dit alles in verschillende condities in het leerproces in termen van leeftijd, blootstelling, geslacht, en opleiding.

Benadering

De bevindingen zijn gebaseerd op een benadering die taaltoetsdata, taaltypologie, en multi-level regressie combineert. De spreekvaardigheid werd nauwgezet beoordeeld. Beoordelaars zijn experts in hun vakgebied, bijvoorbeeld door speciale trainingen. Door middel van vragenlijsten kennen we de beste T2s van de leerders. Door zelfrapportage zal de T2 vaardigheid variëren. Hierdoor kan het T2 effect onderschat worden. De steekproef aan T1s en T2s van de deelnemers in STEX is door de willekeurigheid van immigratiestromen bepaald. Om te controleren voor afhankelijkheden tussen de talen in de steekproef heb ik de richtingscoëfficiënten (slopes) voor het effect van taalafstand afhankelijk van de T1s en T2s laten variëren. De lexicale, morfologische, en fonologische afstandseffecten zijn complementair. Hoewel ze onderling correleren, leveren ze alle drie een eigen bijdrage aan de verklaring van variatie in spreekvaardigheidsscores.

Belang

Leerders met dezelfde taalachtergrond beginnen vanuit een vergelijkbare leertoestand. Leerders zijn in staat om gebruik te maken van hun T1s en T2s om een T3 te leren. Een T2 naast een T1 is nuttig voor het leren van een T3, maar het nut hangt af van de taalafstand.

Daarnaast spelen factoren zoals leeftijd, blootstelling, geslacht, en onderwijs een rol, hoewel deze ondergeschikt zijn aan de rol van taalafstand. Taalafstand belemmert ook de T2 en T3 leerbaarheid van het Nederlands. Leerders worden met deze belemmeringen geconfronteerd. De cognitieve processen die het mogelijk maken om talen te leren zullen meer beslag op de hersencapaciteit leggen als er een grotere taalafstand moet worden overbrugd. Taalafstand verdient het om een prominente rol te spelen in theorieën over leerbaarheid.

Conclusies

Het onderzoek in dit proefschrift maakt gebruik van taaltypologische afstandsmaten voor onderzoek naar tweede taalverwerving. Er is een aantal nieuwe geformaliseerde maten van taalafstand ontwikkeld en getest. De maten laten zien dat taalafstand op lexicaal (woorden), morfologische (zinnen) en fonologisch (klanken) vlak van invloed zijn op de leerbaarheid. Elke maat levert een eigen bijdrage. Taalafstand kan gebruikt worden om de begintoestand van leerders in te schatten, alsook de inspanning die ze moeten leveren om een additionele taal te leren.

Maatschappelijke Relevantie

Het kunnen spreken van meerdere talen is van rechtstreeks nut in het dagelijks leven alsook in het bedrijfsleven. Ongeveer de helft van de mensen in de EU is dan ook meertalig. Alleen al de Europese taalindustrie is een van de snelst groeiende markten, met een omzet van ongeveer 15 miljard Euro (Rehm & Uszkoreit, 2013). Mensen spenderen veel tijd en moeite aan het leren van nieuwe talen. De resultaten uit dit onderzoek laten zien dat het leren van een nieuwe taal afhangt van taalafstand. Landen zoals Denemarken of Zweden betalen taalcursussen van omgerekend ongeveer 500 uur voor iedereen. De cursisten in een klas hebben vaak een gemixte taalachtergrond. Verder onderzoek zal moeten uitwijzen of toegesneden cursussen niet effectiever zijn dan de algemene cursussen die alom aan volwassen leerders worden aangeboden.

CV

Job Schepens studied artificial intelligence at the Radboud University Nijmegen, the Netherlands, where he obtained a BSc in 2008 and an MSc in 2010. His MSc thesis compared distributions of form, meaning, and frequency similarity across languages.

This dissertation results from doctoral research primarily conducted at the Centre for Language Studies (CLS) of the Radboud University.

After a temporary post-doc at the CLS in 2015, Job will start a postdoctoral fellowship awarded by the Max Planck International Research Network on Aging at the Max Planck Institute for Human Development in Berlin, where he will investigate cognitive aging effects on making inferences from experience.