LUND UNIVERSITY

Snakes and Ladders

Developmental Aspects of Lexical-Conceptual Relationships in the Multilingual Mental Lexicon

Suhonen, Lari-Valtteri

2020

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Snakes and Ladders

Developmental Aspects of Lexical-Conceptual Relationships in the Multilingual Mental Lexicon

LARI-VALTTERI SUHONEN
CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

The Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-89213-32-6

Snakes and Ladders

Developmental Aspects of Lexical-Conceptual Relationships in the Multilingual Mental Lexicon

Lari-Valtteri Suhonen

LUND UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Humanities and Theology,
Lund University, Sweden.

To be defended at LUX:C121 on the 12th of December, 2020 at 13:15.

Faculty opponent
Jeanine Treffers-Daller
University of Reading

Organization
LUND UNIVERSITY

Document name:
DOCTORAL DISSERTATION

Date of issue: 12.12.2020

Author: Lari-Valtteri Suhonen

Sponsoring organization Lund University

Title and subtitle
Snakes and Ladders: Developmental Aspects of Lexical-Conceptual Relationships in the Multilingual Mental Lexicon

Abstract

One phenomenon causing issues for language learners in the form of cross-linguistic influence (CLI) is *translation ambiguity* (Eddington & Tokowicz, 2013). Translation ambiguity refers to a situation where word meanings are different in a speaker's languages. To give an example, Swedish does not lexicalize any difference between TO LEND and TO BORROW, whereas this distinction is made in English. Jiang (2002) proposed that language learners depend on explicit rules to resolve translation ambiguity. That is, based on Jiang's predictions, a Swedish learner of English would have to consciously remember this difference to use the two English words successfully. Research in this area has focused on speakers with two languages. This thesis extends the research into third language acquisition.

In this thesis, four empirical investigations are presented. Studies 1 and 2 focus on the *initial state* in L2 and L3 learners, respectively, of a Finnish-based pseudolanguage Kontu. Study 3 explored L1 German and L2 English naturalistic learners of L3 Swedish with longitudinal data from a beginner's level until advanced fluency in the L3. Study 4 is a cross-sectional replication of Study 3. The present thesis represents a unique constellation of studies on CLI in late foreign language learners' *multilingual mental lexicon* (MML) in that it presents data covering the very initial state all the way up to a high (≥ CEFR C1) proficiency. Moreover, it presents data from all six potential directions of CLI in L3 acquisition, in both *accuracy* and *processing*. Finally, all four studies investigated both *forward* and *reverse* CLI in the MML.

Taking the results of the four studies together, CLI in the MML appears to be *multidirectional.* Both forward and reverse CLI was observed. The forward effects align with the predictions of the Parasitic Model (Hall & Ecke, 2003) for the initial stages as well as the RHM-TA overall (Eddington & Tokowicz, 2013). No indications of independence from the previously acquired languages in the L3 lexical representations were found. Also, the results indicate that the effects of translation ambiguity primarily occur in forward CLI at the item level, while the observed effects in reverse CLI were more *global* in nature in line with the predictions of Higby and colleagues (2020). For reverse CLI, there were differences between *immersed* and *non-immersed* learners. Furthermore, CLI operates differently in accuracy and processing. That is, a lack of overt effects does not imply the absence of CLI, which corroborates Jiang's hypothesis. Finally, *cognitive control*, *working memory*, and *psychotypology* were all found to affect the learners' behavior.

The findings highlight the importance of considering the lack of conceptual non-equivalence in modeling multilingual lexical processing as well the importance of separating the effects of *attrition* from the effects of reverse CLI.

Keywords: third language acquisition, cross-linguistic influence, mental lexicon, multilingualism, psychotypology, L3, attrition, second language acquisition, SLA, TLA, transfer, concept, cognitive control, working memory, CLI, multidirectional, multilingual mental lexicon, MML, acquisition, learning, translation ambiguity, introspection, longitudinal

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

ISSN and key title

ISBN: 978-91-89213-32-6

Recipient's notes

Number of pages: 286

Price

Security classification

Snakes and Ladders

Developmental Aspects of Lexical-Conceptual Relationships in the Multilingual Mental Lexicon

Lari-Valtteri Suhonen

LUND UNIVERSITY

The Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-89213-32-6 (print)
ISBN 978-91-89213-33-3 (digital)

Printed in Sweden by Media-Tryck, Lund University, Lund 2020

MADE IN SWEDEN

Media-Tryck is an environmentally certified and ISO 14001 certified provider of printed material.
Read more about our environmental work at www.mediatryck.lu.se

To my grandmother

Table of Contents

Acknowledgement

List of tables

List of figures

Abbreviations

1 Introduction ... 1

1.1 Where do we begin? ... 1

1.2 The aim of this thesis ... 3

1.3 Structure ... 4

2 Background ... 5

2.1 Cross-linguistic influence ... 5

2.1.1 Predicting cross-linguistic influence ... 10

2.1.2 Basic assumptions and individual variation ... 14

2.2 Multilingual mental lexicon ... 19

2.2.1 Building blocks of nominal meaning ... 21

2.2.2 The developing multilingual mental lexicon ... 25

2.3 Previous research ... 34

2.3.1 Resolving translation ambiguity ... 34

2.3.2 Effects of subsequent lexical language acquisition ... 39

2.4 Summary and empirical studies ... 42

3 Cross-Linguistic Influence in Early L2 Word Learning ... 45

3.1 Introduction ... 45

3.2 Aim and predictions ... 47

3.3 Method ... 50

3.3.1 Participants ... 50

3.3.2 Instruments ... 51

3.3.3 Procedure ... 59

3.3.4 Data analysis, transformations, and structure ... 62

3.4 Results 63
3.4.1 Forward CLI in L2 Kontu 63
3.4.2 Reverse CLI in L1 English 73
3.5 Discussion 78
3.6 Conclusion 80
4 Cross-Linguistic Influence in Early L3 Word Learning 81
4.1 Introduction 81
4.2 Aim and predictions 83
4.3 Method 87
4.3.1 Participants 87
4.3.2 Instruments 87
4.3.3 Procedure 92
4.3.4 Data analysis, transformations, and structure 95
4.4 Results 96
4.4.1 Forward CLI in L3 Kontu 96
4.4.2 Reverse CLI in L1 Swedish and L2 English 105
4.4.3 Psychotypology 109
4.5 Discussion 111
4.6 Conclusion 114
5 Longitudinal Aspects of Naturalistic L3 Lexical Acquisition 117
5.1 Introduction 117
5.2 Aim and predictions 119
5.3 Method 122
5.3.1 Participants 122
5.3.2 Instruments 123
5.3.3 Procedure 127
5.3.4 Data analysis, transformations, and structure 131
5.4 Results 131
5.4.1 Forward CLI 131
5.4.2 Reverse CLI 137
5.4.3 Multidirectional CLI in L2 English 142
5.5 Discussion 145
5.6 Conclusion 148
6 Cross-Sectional Aspects of Naturalistic L3 Lexical Acquisition 149

6.1 Introduction ... 149
6.2 Aim and predictions ... 151
6.3 Method ... 153
6.3.1 Participants ... 153
6.3.2 Instruments ... 154
6.3.3 Procedure ... 155
6.3.4 Data analysis, transformations, and structure ... 157
6.4 Results ... 158
6.4.1 Forward CLI in L3 Swedish ... 158
6.4.2 Multidirectional CLI in L2 English ... 161
6.4.3 Reverse CLI in L1 German ... 164
6.5 Discussion ... 167
6.6 Conclusion ... 169
7 Discussion ... 171
7.1 General remarks ... 171
7.2 Direction of CLI ... 175
7.2.1 Forward and reverse CLI ... 175
7.2.2 Introspection ... 182
7.3 Modulating factors ... 184
7.3.1 Proficiency ... 184
7.3.2 Aptitude ... 186
7.3.3 Psychotypology ... 188
7.4 Suggestions for further research ... 192
8 Conclusion ... 195
References ... 197
Appendices ... 213

Acknowledgement

I have been blessed with two marvelous supervisors, Henrik Gyllstad and Tanja Kupisch, who have tirelessly provided much-needed guidance, support, and reinforcement way and beyond expectations. Words cannot express my gratitude.

I am thankful to Malin Bein, Gareth Carrol, Kathy Conklin, Pia Gustafsson, Anna-Lena Jansåker, Scott Jarvis, Fredrik Persson, Anna Ransheim, Jeanine Treffers-Daller, and John Williams for facilitating research visits and data collection. I would also like to thank the mock opponent Kathy Conklin for helpful, constructive feedback in the preparation of the final draft. Furthermore, I would like to thank the faculty opponent and committee for accepting the task.

I have also formed many connections for life. I have my EuroSLA-family Sergio Adrada Rafael, Pernelle Lorette, Verena Platzgummer, Aaricia Ponnet, and Martje Wijers. Spending time with you at conferences (and karaoke) is always one of the highlights of my year. I have my North American colleagues Jennifer Cabrelli, Kara Fleming, Eve Higby, Belem López, Will Travers, and Erica Verde whose impact stretches to the "old continent." I got to study with an amazing group of young MultiMind-researchers. The future is yours. Finally, I have my Konstanz siblings Marieke Einfeldt, Miriam Geiß, Anika Lloyd-Smith, and Sergio Soares. It has been way too long since the last Constanzer Weissbier!

Over the years, I have had the pleasure to work with fantastic colleagues at SOL – as well as supportive fellow doctoral students Juhan Björn, Sandra Cronhamn, Sara Farschci, Matteo Fuoli, Elin Nylander, and Nele Põldvere. Members of the Language Acquisition Seminar have been a great inspiration. Lara Langensee has been an immense help with untangling all things German. I also got to work with Richard Croneberg and Tanya Kolyaka at LDK, and many colleagues at LUS (Bzzz & Finlandsfärjan-forever) and SFS. I should also thank my students, friends, and family. My students constantly remind me about what matters (you will become inspiring teachers!). Also, a big <3 to my friends and family.

The work presented in this dissertation has been (very) generously supported by Lund University, University of Konstanz, The Olof Sager Foundation, Stiftelsen Lundborgska Idofonden, Syskonen Anna Cecilia och Otto Sigfrid Granmarks stipendiefond, Stiftelsen fil. dr. Uno Otterstedts fond för främjande av vetenskaplig undervisning och forskning, and Elisabeth Rausings Minnesfond.

Finally, I would like to thank the almost two hundred unique participants in the four main experiments, pilots, and stimulus norming studies. Without you, none of the research could have been done.

Any shortcomings, if present, are my own.

List of tables

Table 1: Dimensions of cross-linguistic influence ... 6
Table 2: Explanations for cross-linguistic influence ... 9
Table 3: Stages in the Parasitic Model ... 32
Table 4: Modulating factors in the Parasitic Model ... 33
Table 5: Controlled, measured, and manipulated factors ... 43
Table 6: First stage of the Parasitic Model ... 48
Table 7: Form-meaning mappings in Kontu v1 ... 52
Table 8: Form-meaning mappings in Kontu v1 ... 53
Table 9: Forward CLI in learning (processing) ... 65
Table 10: Time on task in assessment ... 66
Table 11: Forward CLI in assessment (processing) ... 68
Table 12: Sleep quality and retention ... 70
Table 13: Forward CLI in assessment (accuracy) ... 72
Table 14: Reverse CLI (priming) ... 77
Table 15: First stage of the Parasitic Model ... 84
Table 16: Psychotypology (I-CLSA stimuli) ... 91
Table 17: Forward CLI in learning (processing) ... 99
Table 18: Forward CLI in assessment (processing) ... 100
Table 19: Forward CLI in learning (processing) ... 101
Table 20: Forward CLI in assessment (accuracy) ... 104
Table 21: Reverse CLI (priming) ... 108
Table 22: Psychotypology (E-CLSA evaluations) ... 110
Table 23: Psychotypology (I-CLSA performance) ... 110
Table 24: Self-assessment scale of language competence ... 124
Table 25: Word Pair Similarity Perception Task ... 125
Table 26: Tasks and instruments at different data points ... 130
Table 27: Forward CLI (similarity ratings) ... 135
Table 28: Forward CLI (ToTs) ... 136
Table 29: Reverse CLI (similarity ratings) ... 140
Table 30: Reverse CLI (ToTs) ... 141
Table 31: Multidirectional CLI (similarity ratings) ... 143
Table 32: Multidirectional CLI (ToTs) ... 144
Table 33: Forward CLI in L3 Swedish (similarity ratings) ... 159
Table 34: Forward CLI in L3 Swedish (ToTs) ... 160
Table 35: Multidirectional CLI in L2 English (similarity ratings) ... 162
Table 36: Multidirectional CLI in L2 English (ToTs) ... 163

Table 37: Reverse CLI in L1 German (similarity ratings) 165
Table 38: Reverse CLI in L1 German (ToTs) 166
Table 39: Controlled, measured, and manipulated factors 174
Table 40: Directions of CLI in the present thesis 175
Table 41: Overview of forward CLI 177
Table 42: Overview of reverse CLI 179
Table 43: Correlation of aptitude measures 187
Table 44: Overview of aptitude 188
Table D1: Prime-target pairings in the distractor condition 219
Table D2: Prime-target pairings in the 'kontu' condition 219
Table D3: Prime-target pairings in the baseline condition 220
Table D4: Prime-target pairings in the 'pseudo' condition 221
Table E1: Acquiring form-meaning mappings in Kontu v1: ToTs 223
Table E2: Distribution of RT data in priming tasks 225
Table E3: Self-ratings of aptitude 226
Table E4: Performance in the flanker-task 226
Table E5: Working memory scores 227
Table E6: Correlation of aptitude measures 228
Table E7: ToTs for learning sequences 230
Table E8: First instance vs. second instance (LME): learning 231
Table E9: First instance vs. second instance (LME): assessment 232
Table E10: RTs in Reverse CLI 233
Table G1: Prime-target pairings in the distractor condition 236
Table G2: Prime-target pairings in the 'kontu' condition 236
Table G3: Prime-target pairings in the baseline condition 237
Table G4: Prime-target pairings in the 'pseudo' condition 239
Table I1: Forward CLI in learning (processing) 242
Table I2: Reverse CLI (priming) 243
Table I3: Self-ratings of aptitude 245
Table I4: Cognitive control 245
Table I5: Working memory 246
Table I6: Forward CLI in learning (processing) 247
Table I7: RTs in priming 248
Table N1: Correlation of proficiency measures in Swedish 267

List of figures

Figure 1. The lexical entry in the mental lexicon 19
Figure 2. A three-stage system for vocabulary acquisition in SLA 26
Figure 3. The Revised Hierarchical Model 28
Figure 4. The Distributed Feature Model.. 29
Figure 5. The Modified Hierarchical Model.. 30
Figure 6. The Revised Hierarchical Model of Translation Ambiguity............. 35
Figure 7. Learning sequences and blocks in Study 1 53
Figure 8. The learning outcome test 54
Figure 9. Sequence of presentation in the priming task 56
Figure 10. Stimulus in the n-back task of working memory 57
Figure 11. Stimulus in the flanker task 58
Figure 12. Procedure in Study 1 60
Figure 13. Time spent on task (ms) in learning sequences 64
Figure 14. Time spent on task (ms) in LOTs 67
Figure 15. Accuracy in assessment modules 69
Figure 16. Accuracy in the masked priming task 73
Figure 17. Visualization of RT data.. 74
Figure 18. Visualization of data for assessing learning effects. 75
Figure 19. Visualization of data for assessing learning effects 76
Figure 20. Stimulus presentation in E-CLSA.. 89
Figure 21. Stimulus presentation in I-CLSA... 92
Figure 22. Procedure in Study 2 93
Figure 23. Time spent on task (ms) in learning sequences 97
Figure 24. Accuracy in assessment modules 102
Figure 25. Accuracy in pre- and post-test conceptual priming tasks 105
Figure 26. Visualization of RTs by condition and time 106
Figure 27. Visualization of data 107
Figure 28. Experimental instruments used in Study 3 123
Figure 29. Word Pair Similarity Perception (WSPT) task item example........ 126
Figure 30. Procedure in Study 3 129
Figure 31. WSPT similarity ratings in forward CLI...................................... 132
Figure 32. WSPT time on task in forward CLI 133
Figure 33. WSPT similarity ratings in reverse CLI 138
Figure 34. WSPT time on task in reverse CLI 139
Figure 35. Experimental instruments in Study 4 154
Figure 36. Procedure in Study 4 156

Figure 37. Marginal effects of WSPT ratings in Swedish 158
Figure 38. Marginal effects of WSPT ratings in English 161
Figure 39. Marginal effects of WSPT ratings in German 164
Figure 40. Overview of Studies 1-4 173

Abbreviations

AO	Age of onset
ATH	Activation Threshold Hypothesis
BBM	Bit-by-Bit Model
CC	Cognitive control
CLA	Cross-language activation
CLI	Cross-linguistic influence
L1	First language
L2	Second language
L3	Third language
Ln	Additional language
LTM	Long-term memory
LME	Linear mixed effect model
LoE	Length of exposure
LOT	Learning outcome test
LPM	Linguistic Proximity Model
MHM	Modified Hierarchical Model
MML	Multilingual mental lexicon
ms	Millisecond
PM	Parasitic Model
RHM	Revised Hierarchical Model
RHM-TA	Revised Hierarchical Model of Translation Ambiguity
RIF	Retrieval-induced forgetting
RT	Response time
SL	Source language in CLI
SLA	Second language acquisition
SM	Scalpel Model
STM	Short-term memory
TL	Target language in CLI
TLA	Third language acquisition
ToT	Time on task
TPM	Typological Primacy Model
WM	Working memory

1 Introduction

1.1 Where do we begin?

Already in 1953, Weinreich suggested that interaction of languages in a multilingual speaker is an unavoidable feature of language learning and use. More recently, in the case of vocabulary, Bardel (2015) points out that the character of the interaction of languages in a multilingual speaker changes as a result of development. In the beginning, overt effects – such as instances of switching to another language – dominate, while when proficiency increases, the effects are often related to aspects of word meaning rather than form. Ringbom (2007) similarly suggests that meaning-based *cross-linguistic influence* is more prevalent in advanced learners.

This dissertation is about cross-linguistic influence in the developing *multilingual mental lexicon* (MML) in adults. Cross-linguistic influence (CLI) is "the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" (Odlin, 2003, p. 436). As this thesis investigates all possible directions of CLI in multilingual speakers, a *target language* (TL) can be any language in which CLI is being observed – irrespective of order of acquisition. The language, from which CLI takes place, is referred to as a *source language* (SL). This aligns with the purpose of research on *third language* (L3) acquisition, which considers the "complex constellations of languages that occur with multilingual speakers and explor[es] the roles of these languages in the acquisition process" (Hammarberg, 2018, p. 127). The observed languages in this dissertation are German, English, Swedish, and Kontu which is an artificial language based on Finnish.

Potential effects of CLI will be investigated with regard to the *mental lexicon* which following Jarema and Libben (2007) is "the cognitive system that constitutes the capacity for conscious and unconscious lexical activity" (p. 3). Unconscious lexical activity encompasses the covert processes that due to the time frame (first few hundred milliseconds of activation) allow little to no conscious intervention (p. 3). The primary focus of this dissertation is on aspects of development and directionality, i.e., to what extent the different languages influence each other at different stages of acquisition in relation to conscious and unconscious lexical activity. To this end, four empirical studies have been carried out.

Recently, several models of third language acquisition[1] have been proposed, including the Scalpel Model (Slabakova, 2016), the Typological Primacy Model (Rothman, 2015), L2 Status Factor (Bardel & Falk, 2007, 2012; Falk & Bardel, 2011), the Cumulative Enhancement Model (Flynn, Foley & Vinniskaya, 2004), the Linguistic Proximity Model (Westergaard, Mitrofanova, Mykhaylyk & Rodina, 2016), and the Bit-by-Bit Model (Domene Moreno, 2019). These models, however, do not make predictions about the development of the lexicon (as opposed to morphosyntax and phonology) in *third language acquisition* (TLA). To my knowledge, the only model of TLA that makes specific predictions about the development of vocabulary is Hall and Ecke's (2003) *Parasitic Model* (PM). Since the model makes developmental predictions for MML in the L3, Studies 1 and 2 in this dissertation take this model as their point of departure.

CLI in the lexicon in TLA has been extensively researched in *oral production* (e.g., Bardel, Gudmundson & Lindqvist, 2012; Cenoz, 2001, 2003; De Angelis & Selinker, 2001; Dewaele, 1998; Hall & Ecke, 2003; Hammarberg, 2001, 2009; Lindqvist, 2009, 2010; Lindqvist & Falk, 2014; Neuser, 2017; Singleton, 1987) and *writing* (e.g., De Angelis, 2005a, 2005b; Ecke, 2001; Jarvis, 1998; Neuser, 2017; Ringbom 1987, 2001; Singleton & ó Laoire, 2006). While these studies provide insight into the organization of the MML, they have relatively little to say about the underlying changes in representation that potentially lead to overt, noticeable changes in production. Gradual changes in underlying representation do not necessarily manifest themselves in production. In addition to measures of accuracy, all four experiments in this thesis have been designed to tap into the speaker's unconscious representations and processing, and how these change over time.

In relation to multilingual linguistic development in general, four major factors have been proposed to affect the source, quality, and quantity of CLI: *proficiency*, *recency*, *psychotypology*, and *L2 status* (Boratynska-Sumara, 2014; Dentler, 2000; Neuser, 2017; Williams & Hammarberg, 1998). Out of these, proficiency and psychotypology lend themselves to be measured. Proficiency and recency are interdependent factors and recency in is subsumed as a component of proficiency in this thesis. L2 status (see 2.1.1) refers to a *foreign language effect*. Psychotypology

[1] In this dissertation, *acquisition* is used to refer to language learning irrespective of age of onset.

refers to *perceived* language distance (Kellerman, 1983). Furthermore, *cognitive control* (e.g., Green, 1998) and *working memory* (e.g., Papagno & Vallar, 1992; Service & Kohonen, 1995) have been postulated to have a major impact on the acquisition and processing of vocabulary (Ellis & Sinclair, 1996). In this thesis, cognitive control and working memory are subsumed under *aptitude*. The three measures that have been included as predictors in the empirical studies in this dissertation are proficiency, aptitude, and psychotypology.

1.2 The aim of this thesis

The over-arching aim of this dissertation is to investigate the developing multilingual mental lexicon (MML) focusing particularly on cross-linguistic influence (CLI). Given this aim, the following three research questions have been formulated:

RQ1 Is cross-linguistic influence (if present) in the multilingual mental lexicon unidirectional or multidirectional?

RQ2 To what extent is cross-linguistic influence in the multilingual mental lexicon affected by proficiency, aptitude, and psychotypology?

RQ3 Are the aforementioned aspects of directionality and the effect of modifying factors dependent on the stage or type of acquisition?

The first research question relates to the direction of CLI. All potential directions between the speaker's languages are of interest. The second research question relates to factors that represent individual variation in language learners that have been identified in previous research to affect language acquisition. Proficiency, in this thesis, includes the following subcomponents: use, recency, age of onset, length of exposure, and manner of acquisition. Aptitude is operationalized as working memory and cognitive control. Psychotypology is treated both as a conscious and unconscious construct in that it is measured both overtly and covertly. The third research question relates to the effect of stage of acquisition. This dissertation takes a semi-longitudinal approach in that all studies either represent data from multiple points in the learning process from the same set of participants or use a cross-sectional design with development being operationalized as proficiency.

1.3 Structure

This dissertation consists of eight chapters of which the first chapter is the present introductory chapter. Chapter 2 is devoted to theoretical background. Section 2.1 covers CLI from the perspective of predicting it. Individual variation is covered in subsection 2.1.2. Section 2.2 presents developmental aspects of the MML, section 2.3 presents previous, relevant, research in the area, and section 2.4 provides a brief summary of the background chapter.

The empirical work in this dissertation consists of four studies. Chapter 3 presents a study on CLI in early L2 word learning (with English as the L1 and an artificial language 'Kontu' as the L2) and Chapter 4 presents a study on CLI in early L3 word learning (with Swedish L1, English L2, and Kontu as the L3). Chapter 5 presents a longitudinal study on CLI in naturalistic L3 lexical acquisition (with German L1, English L2, and Swedish as the L3), and Chapter 6 a cross-sectional study on CLI in naturalistic L3 lexical acquisition (with German L1, English L2, and Swedish as the L3).

Chapter 7 combines and discusses the results of the four empirical studies in relation to the overarching research questions. Finally, Chapter 8 contains a brief summary of the results of the four studies.

Note. The first part of the title of this thesis alludes to the board game Snakes and Ladders. It is meant to illustrate how as learners with an already established mental lexicon, we can learn new words by attaching new forms to existing form-meaning mappings. In many cases, this leads to fast and efficient learning (i.e., the ladders). However, more often than one might think, word meaning is variable across languages, causing issues in acquisition and processing (i.e., the snakes).

2 Background

2.1 Cross-linguistic influence

Psychology and linguistics have, for many decades, attempted to explain and document differences between native and non-native language acquisition. It is uncontroversial to assume that previous linguistic experience affects language acquisition. We refer to this effect broadly as *cross-linguistic influence* (CLI). As mentioned earlier, this dissertation uses Odlin's (2003) definition that CLI is "the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" as a point of departure (p. 436). In relation to the overarching research questions in this thesis, there are three aspects of this definition that should be pointed out. Firstly, the definition includes both similarities and differences, which often lead to positive and negative results in learning. Secondly, the definition implies that linguistic systems of language learners, while systematic, are often not native-like. Finally, it does not take any stance with respect to the direction of CLI.

Several terms for CLI have been used throughout history such as *transfer* (Thorndike, 1923) and *interference* (Weinreich, 1953). Transfer, in the early sense, focused primarily on positive effects of linguistic similarity. Interference is seldomly used in recent research and refers primarily to the negative aspects of CLI (Odlin, 1989). In the 1980s some researchers felt that transfer was not an appropriate term to describe the phenomenon since it was associated with the behaviorist notion of skill transfer (Odlin, 1989; Jarvis & Pavlenko, 2008). Sharwood-Smith and Kellerman (1986) proposed CLI as a theory-neutral term to describe a broad range of influences. Unlike approaches assuming behaviorist skill transfer, some recent language acquisition literature takes *cross-linguistic influence* to include changes in performance whereas *transfer* primarily refers to changes in representation (Lloyd-Smith, 2020; Rothman et al., 2019).

Jarvis and Pavlenko (2008) point out that CLI has often been seen as something negative and that the focus has been on the effect of the mother tongue on subsequent languages focusing on the societal majority language. The terms *positive transfer* and *negative transfer* are sometimes used to differentiate between instances where CLI is advantageous to the learner from the ones where it is disadvantageous (Odlin, 2012). Positive transfer could, for example, be successful use of *lexical inferencing*[2] from German to understand the meaning of the Swedish near-cognate 'fönster' WINDOW[3] based on German 'Fenster.' An example of negative transfer would be the use of Swedish 'snäll' KIND by a speaker of German to refer to something moving in a fast manner due to the German false friend 'schnell' FAST. These examples refer to CLI being positive or negative in a specific use situation rather than as a generic strategy.[4]

CLI can manifest itself in many different dimensions of language. Virtually any area of language knowledge and use can be subject to CLI. Table 1 below highlights different potential areas of CLI.

Table 1

Types of cross-linguistic influence associated with dimensions of language (Jarvis & Pavlenko, 2008)

Dimension	Types of transfer
1. Area of language knowledge	phonological, orthographic, lexical, semantic, morphological, syntactic, discursive, pragmatic, sociolinguistic
2. Directionality	forward, reverse, lateral, bi- or multidirectional
3. Cognitive level	linguistic, conceptual
4. Type of knowledge	implicit, explicit
5. Intentionality	intentional, unintentional
6. Mode	productive, receptive
7. Channel	aural, visual
8. Form	verbal, nonverbal
9. Manifestation	overt, covert
10. Outcome	positive, negative

[2] *Lexical inferencing* refers to the processes involved in "making informed guesses as to the meaning of a word in light of all available linguistic cues in combination with the learner's general knowledge of the world, [..] awareness of the co-text and [..] relevant linguistic knowledge" (Haastrup, 1991, p.13).

[3] In this dissertation, word forms are marked with single quotation marks i.e., 'fönster.' Meaning is marked with small caps i.e., WINDOW.

[4] *Generic strategy* here refers to varying levels of inhibition of a particular source language by default.

An important topic of interest in modeling CLI is its *direction* (see Table 1 above). Effects of subsequent language(s) on the speaker's existing language(s) have not been researched extensively until recent years but the notion is not new. Mencken (1937) described both lexical and syntactic CLI in the mother tongues of immigrants to the United States. This type of CLI whereby a later acquired language influences the mother tongue is often referred to as *reverse transfer* or *regressive transfer*, as opposed to *forward transfer* where the mother tongue influences a later acquired language. There are four logical possibilities for forward transfer from the previously acquired two languages to an additional language: 1) no transfer, 2) only transfer from the L1, 3) only transfer from the L2, and 4) transfer from the L1 and the L2 (Rothman, 2015). If the language learner speaks multiple languages beyond the mother tongue, there is a possibility of *lateral transfer* where the second language influences the third language (Jarvis & Pavlenko, 2008). All the aforementioned types of CLI can be said to be unidirectional, with influence taking place in one specific direction. Another theoretical possibility is that CLI is *multidirectional* in nature, meaning that all languages, or constructions herein, influence each other with varying quantity and quality (Sharwood-Smith, 1989). The empirical studies in this dissertation adhere to the assumption (Jarvis & Pavlenko, 2008) that there is at least a theoretical possibility that all languages of a speaker influence each other.

Several explanations for the existence of CLI have been proposed. For example, Odlin (1989) focuses on the aspect of languages affecting each other, Selinker (1992) on CLI being a constraint in learning, and Krashen (1983) on CLI being a gap-filling strategy in the acquisition process. As for the explanations to how CLI takes place, the following accounts have been proposed: 1) *spreading activation* and *non-selective access*, 2) *attrition*, 3) *retrieval-induced forgetting* (RIF), and 4) *inferencing*. These are presented in detail below.

Based on Collins and Loftus (1975), it has been proposed that *spreading activation* causes unavoidable automatic activation of related properties, particularly semantic content (Anderson, 1983). The *Associative Networks Theory* proposes that the semantic memory – the store for all types of factual knowledge – is a network of associated ideas and concepts. When a concept (e.g., FIRE ENGINE) is activated in this network, related concepts (e.g., RED) are also partially activated through spreading activation (Collins & Loftus, 1975).

Non-selective access of language refers to involuntary activation of the speaker's all languages (Kroll, Bogulski & McClain, 2012; Tokowicz, 2014). This means that if a speaker activates the Finnish 'sitruuna' LEMON, then the translation equivalents of 'lemon' in all the speaker's languages are activated. Both spreading activation and non-selective access refer to an instance of *priming*, which is the activation of one structure or concept by another. As terms, spreading activation is often used for within-language co-activation while non-selective access is used for between-language co-activation, even if the underlying mechanism is the same.

Explanations of gradual changes in representation include *attrition* and *retrieval-induced forgetting*. Schmid and Köpke (2017) suggest that acquisition and attrition[5] are interrelated in that they are two sides of the same process. Gyllstad and Suhonen (2017) suggest that this aligns with the notion of *retrieval-induced forgetting*, where each instance of recalling a particular item causes suppression of its related items. RIF occurs when recalling a memory trace causes loss of related information and is caused by closely related lexical items competing with each other (Anderson, Björk & Björk, 1994). The difference here is that attrition, by default, takes place due to lack of activation of a particular memory trace whereas RIF is caused by activation of associated traces. All of the aforementioned processes are beyond an individual's conscious control.

CLI as an intentional, conscious communicative strategy, *inferencing*, has been a rather unexplored research area of CLI research with most studies focusing on negative aspects of CLI even if it can be postulated that CLI has more positive impact on post-L1 acquisition (Jarvis & Pavlenko, 2008; Ringbom, 2001, 2007). Recent work has, however, tackled this question. The use of CLI as an intentional strategy has been highlighted by Fuster and Neuser (2020), and Neuser (2017). This can include, for example, inferencing and the success rate depends on the rate of linguistic and conceptual similarity of the items (and languages) in question. Failed attempts of lexical inferencing, for example due to false friends, are likely explanations for many observed instances of overt CLI.

The aforementioned proposals for the occurrence of CLI are not necessarily contradictory. They explain different types of CLI and vary in their scope. There are two aspects that should be highlighted with respect to these: whether the explanations relate to a process that is *instantaneous* or *gradual* and whether the process is intentional. Instantaneous here refers to something taking place automatically and very rapidly. This is typically referred to as *priming* in processing research. Gradual, on the other hand, refers to development that takes place over time.

[5] The term *attrition* has typically very negative connotations and a strong conventionalized meaning referring to clinical cases of language loss.

Table 2 below presents the aforementioned explanations for the occurrence of CLI based on these aspects.

Table 2

Explanations for cross-linguistic influence

	UNINTENTIONAL	INTENTIONAL
PRIMING	Spreading activation Non-selective access	Inferencing
GRADUAL	Attrition Retrieval-induced forgetting	

Note. Priming refers to the instantaneous activation of one structure or concept by another. Spreading activation and non-selective access are both processes that take place automatically at a given language use event. Attrition and retrieval-induced forgetting are unavoidable processes that take place gradually over time. Inferencing is a conscious strategy to attempt to make informed guesses about meaning using existing knowledge. While inferencing is used in a given moment and often in a very automated way, it is a skill that has a developmental aspect to it.

As this dissertation focuses on developmental aspects, its focus is on those processes that are subject to long-term representational and processing changes, i.e., not related to temporary lapses in production. Given that aim, the focus is on conceptual, unintentional, covert changes (see Table 1 above) in both positive and negative aspects of CLI in all directions of influence in lexical development. As such, we need to make a distinction between CLI that is *linguistic*, i.e., relates to linguistic forms and structures, and *conceptual*, i.e., relates to the mental concepts underlying these forms and structures (Jarvis & Pavlenko, 2008).

Furthermore, in this dissertation the term CLI is used to refer to both between-language effects at a particular time point as well as gradual changes in representation caused by similarity or the lack thereof between a speaker's languages. The term *transfer*[6] will only be used when employed by cited authors. *Attrition* will only be used to refer to loss of language due to lack of use unless used differently by the cited authors. Furthermore, it is assumed that the source, quality, and quantity of CLI is affected by both intra- and extralinguistic factors. Intra-linguistic factors are covered in 2.1.1 below and individual variation in 2.1.2.

[6] See Rothman, Gonzáles Alonso, and Puig-Mayenco (2019), as well as Odlin and Yu (2016) for a terminological discussion.

2.1.1 Predicting cross-linguistic influence

Several models of multilingual language acquisition have been proposed that focus on predicting CLI based on one of the following three aspects: *cognitive similarity*, *typological distance*, and *structural similarity*. Cognitive similarity refers to similarity in mental representation and processing that is caused by similar manner of acquisition (Bardel & Falk, 2012). Typological distance refers to similarity of languages, typically derived from genealogical relationship between languages (Rothman, 2015), whereas structural similarity refers to similarity between languages in particular structures within those languages (Slabakova, 2016).

The L2 Status Factor is a model that proposes that for those learners who have acquired their second language in a qualitatively similar manner as the third language, the second language will pose a more likely source for CLI in morphosyntax (Williams & Hammarberg, 1998; Bardel & Falk, 2007, 2012). This is attributed to a "higher degree of *cognitive similarity* [my italics] between the L2 and the L3 than between the L1 and the L3" (Bardel & Falk, 2012, p. 3). Cognitive similarity, here, refers to the storage of the language in the brain modulated by the manner of acquisition. Better metalinguistic awareness of the L2 has also been proposed as one of the explanations of the status of the L2 as the more likely source (Ortega, 2008). The original hypothesis for this was proposed by Williams and Hammarberg (1998), who found that involuntary switches in L3 production reverted primarily to the L2, while the L1 was used for conscious explanations. Previous research (e.g., Ringbom, 1983; Stedje, 1977; Vildomec, 1963) had found that unintentional language switches took place more often with function words rather than content words. Observations of the effects of the L2 on the L3 led to a suggested *foreign language effect* (Meisel, 1983). Williams and Hammarberg (1998), then, hypothesized that a qualitative difference existed between the L1 and the L2 in the process of L3 production. This hypothesis aligns with the procedural–declarative[7] distinction (Paradis, M., 2004, 2009; Ullman, 2001, 2005), where late acquired languages are "sustained to a large extent by declarative memory" (Paradis, M., 2009, p. 173). The hypothesis has been further developed

[7] *Declarative memory* is also sometimes referred to as explicit, with the memory processes taking place with conscious recall. Declarative memory holds, for example, conceptual information. Procedural memory is also sometimes referred to as implicit or nondeclarative. These memory processes take place without conscious recall. This is not to say that we cannot have conscious memories about morphosyntactic rules even as native speakers. The assumption is that on-line language processing of those rules takes place primarily through the procedural system.

into the L2 Status Factor Model (Falk & Bardel, 2010, 2011) with a primary focus on morphosyntax. In terms of vocabulary, there is a presumption that vocabulary is primarily aligned with declarative memory systems. Hence, a similar cognitive difference between the first and the second language in relation to the third would not exist. For that reason, the second language, in the present instantiation of this approach, does not enjoy a special status in the acquisition of third language vocabulary.

Typological distance is another aspect that has been highlighted as an explanatory factor in multilingual language acquisition. Closely related languages have often been found to be more likely coactivated than unrelated languages (Ringbom, 2007, p. 91), and researchers have found that CLI tends to take place from the typologically closer language (e.g., Cenoz, 2001, 2003; De Angelis, 2005; De Angelis & Selinker, 2001; Leung, 2005; Lindqvist, 2009; Lindqvist & Bardel, 2013; Odlin & Jarvis, 2004; Ringbom, 1987; Singleton & Ó Laoire, 2006). Typological distance has not always been well defined (Lindqvist, 2015). It has either been postulated as the *genealogical relationship* of the two languages, or the relationship of particular structures of the languages (Bardel & Falk, 2012). Genealogical relationship is the relationship between languages from the perspective of language families. However, typology in language acquisition can also relate to *psychotypology*, which is *perceived* language distance introduced by Kellerman (1983). Psychotypology has been treated both as a *static* factor (e.g., Flynn, Foley & Vinnitskaya, 2004; Rothman, 2015) and a *fluid* factor (e.g., Neuser, 2017; Rast, 2008; Sayehli, 2013; Suhonen, 2015; Xia, 2017). In the case of the former, psychotypology can be *predicted*, whereas in the latter it has to be *measured*. While in this subsection psychotypology is presented as a static or universal mental representation that is sensitive to linguistic cues, it will be argued in subsection 2.1.2 that psychotypology can also be conceived as a fluid factor that is subject to within-subject and between-subject variation.

Typological distance as an explanatory factor in third language acquisition is highlighted by the *Typological Primacy Model* (TPM) which proposes that a deciding factor for choosing which language acts as a donor language in third language acquisition is *perceived structural similarity* (Rothman, 2015, Rothman et al., 2019). Given that the TPM proposes that it is *perceived* structural similarity that guides the choice of source language, defining similarity is of extreme importance when testing the tenets of TPM. Similarity, here, is defined as what the "internal parser takes to be most similar (actual or perceived) structural similarity" among the three grammars (Rothman, 2015, p. 183). This structural similarity is determined by cues: first lexical similarity, then phonology, morphology, and finally syntax (Rothman, 2015; Rothman, Gonzáles Alonso & Puig-Mayenco, 2019). The TPM is strictly a model of what happens at the *initial state*. The initial stage is defined to be "what the acquirer brings to the first moments of exposure to input" (Rothman, 2015, p. 179). The TPM explains both facilitative and non-facilitative transfer.[8] It is a *best guess* type of model from the learner's perspective and it is a model of unconscious transfer: the learner is not consciously in control of the transfer experience. The proposal is that before the selection of the donor language there is a brief *transitory initial stage* during which the learner has access to both the L1 and the L2 systems, which facilitates the comparative process. Then one of those systems is transferred[9] in a wholesale manner (in its entirety), and all initial hypotheses for the L3 are made based on the transferred features. The TPM rests on the assumption that wholesale transfer is the most economical way for the brain to deal with early stages of the acquisition of the third language. Rothman (2015) suggests that it is "the mind's predisposition to put forth the least amount of effort towards a cognitive task" (p. 180).

Multiple models emphasize the importance of *structural similarity* at a property level in multilingual language acquisition. These include the *Scalpel Model* (Slabakova, 2016), the *Linguistic Proximity Model* (Westergaard, Mitrofanova, Mykhaylyk & Rodina, 2016), and *the Bit-by-Bit Model* (Domene Moreno, 2019). The Scalpel Model (SM) aims to identify "what happens beyond the initial state of acquisition and what factors may influence change from one state of knowledge to another" (Slabakova, 2016, p. 2). It takes neurocognitive models of the multilingual brain as its base, particularly those of Abutalebi and Green (2007) and M. Paradis (2004), and builds upon the assumption that all linguistic knowledge is interconnected and that the languages of the individual are not functionally separate. These assumptions can be connected to the aforementioned language non-selectivity in multilingual language processing. Effects of age of acquisition and proficiency are expected (Slabakova, 2016). Furthermore, the SM incorporates some of the features of the Cumulative Enhancement Model (Flynn, Foley & Vinnitskaya, 2004) as well as the TPM (Rothman, 2015), namely that neither the L1 nor the L2 is a privileged supplier of CLI in L3 acquisition. Wholesale transfer

[8] *Facilitative* and *non-facilitative transfer* can more or less be mapped to the aforementioned terms positive and negative transfer.

[9] It is unclear whether the use of the word 'transfer' here is metaphorical in nature or if the proposal is that the contents of the donor system are literally copied to a separate physical representation.

of one of the previously acquired languages does not take place: it is not necessary (Slabakova, 2016). This is the opposite of the claim of the TPM, where the assumption of wholesale transfer is motivated based on processing economy (Rothman, 2015). According to Slabakova, wholesale transfer might be beneficial in L2 acquisition since the learner can use the L1 as a resource. However, in L3 acquisition, the learner already has two appropriately tagged language systems. Since allowing the influence of one system means blocking the other system, the costly *inhibition* process (i.e., the suppression of task-irrelevant stimuli) has to take place in any case. In that sense, the learner would be better off blocking features one-by-one selecting the features that are most likely to be facilitative (Slabakova, 2016).

Similar to the SM, the Linguistic Proximity Model (LPM) proposes that different learning patterns exist for different properties of language. This can apply to a wide range of such properties such as case marking or individual form-meaning pairings (Westergaard, Mitrofanova, Mykhaylyk & Rodina, 2016). The Bit-by-Bit Model (BBM), which has been developed within the area of phonological acquisition, also assumes that transfer depends on the linguistic property to be acquired, suggesting that phonological grammar is "made up of individual elements called bits" and that language transfer affects each of these bits separately (Domene Moreno, 2019, pp. 40-41). This means that CLI is inherently selective property-wise in SM, LPM, and BBM.

The presented models focus on predicting CLI based on one factor at a particular stage of language acquisition, or in a particular domain of language. The TPM suggests transfer from the typologically closest language. The strength of the TPM is that it makes very specific predictions once the closest language has been determined. It easily lends itself to empirical testing. The TPM, however, only considers the very initial state of L3 acquisition. By contrast, Slabakova (2016) stresses that theories should go beyond the initial state and formulate testable hypotheses regarding the later stages of L3 development. An important addition of the SM is its attempt to model default trajectories for L3 acquisition, to define separate *modulating factors* that influence individual language learners as well as individual language constructs in the learning process of those individuals. Crucially, the models presented so far say little about the development of vocabulary (as opposed to morphosyntax and phonology) in multilingual language acquisition. The SM, the LPM, and the BBM do not make any predictions about developmental trajectory for vocabulary. In the TPM, vocabulary plays a role in that the amount of shared vocabulary between two languages can predict the source of morphosyntactic transfer. It does not, however, make any specific predictions with respect to the source of transfer in lexicon. We could, however, postulate that the same perceived structural similarity assessment could act as a base for predicting from which language vocabulary transfer predominantly would take place at the initial stage. A model that makes specific predictions about the development of vocabulary, albeit only in forward CLI, is Hall and Ecke's (2003) Parasitic Model, which is presented in subsection 2.2.2.

2.1.2 Basic assumptions and individual variation

In addition to universal factors that affect CLI in multilingual language learners, CLI is also subject to individual variation. Research in the past few decades has come to the conclusion that all of a speaker's languages operate in a single language system (Kroll, Dussias, Bice & Perrotti, 2015, see Kroll & Tokowicz, 2005 for an extensive overview of the debate). This argument can be referred to as the *single system assumption*. It is based on findings that 1) language access is *non-selective*, i.e., all languages of a speaker are active in both comprehension and production (e.g., Costa et al., 2000; Dijkstra, 2005; Kroll et al. 2006; Kroll, Bogulski & McClain, 2012; Marian & Spivey, 2003; Tokowicz, 2014; Van Hell & Dijkstra, 2002), and 2) that increasing proficiency in additional languages affects the first language proportionally (e.g., Abutalebi & Green, 2007; Steinhauer et al., 2009). A single system assumption has consequences to what aspects of individual variation are relevant for research on language processing. Green's (1986, 1998) *Inhibitory Control Model* suggests that the natural state for languages is *activation* and regulating languages is performed through *inhibition,* i.e., suppression of task-irrelevant stimuli. The inhibition cost for languages with a higher *proficiency* is larger since activation levels for highly proficient languages are higher than for languages with lower proficiency. One prediction that the model makes is that the switching cost to a more proficient language is higher than to a less proficient language. *Psychotypology*, the perceived language distance between the languages of the learner, matters in estimating what the learner is likely to attempt to inhibit. Thus, the three factors that have been included in the empirical studies in this dissertation are *proficiency*, *aptitude*, and *psychotypology.*

The first factor of individual variation in this dissertation is *proficiency*, which is an aspect that inherently varies from one language learner to the other. In addition to between-subject variation in the proficiency in other previously acquired languages as well as in the tested language, there is within-subject variation in different domains of language knowledge. A particular learner can have a relatively high level of proficiency in, say, speech production but a low level of proficiency in writing. Since multilinguals use different languages for different tasks in everyday life, the default language for different tasks can also vary irrespective of proficiency. Hence, proficiency for individual tasks is also dependent on the typical language used for that given task (Wei, 2000). In all four empirical studies, proficiency is measured using vocabulary size as a proxy. Proficiency in other domains (writing, reading, speaking, listening, as well as use in home, work, school, etc.) of language knowledge is self-reported by the participants through a range of questions in the background questionnaire in all four empirical studies.

Level of proficiency in the L2, as well as the L3, have been shown to have an effect with respect to CLI as studies have found that the amount of CLI decreases as the proficiency in the L2/L3 increases (e.g., Lindqvist, 2009; Navés et al., 2005; Neuser, 2017; Singleton, 1987; Williams & Hammarberg, 1998), although there is evidence for CLI in near-native speakers as well (Abrahamsson & Hyltenstam, 2009; Birdsong, 2006; Lardiere, 2007). Proficiency is central for the four empirical studies in this dissertation as the main novelty of this dissertation is looking at third language lexical acquisition from a developmental perspective. While the decreasing observable effect of the background languages in the language being learned is rather expected, one perhaps more interesting aspect is that not all types of CLI appear at all stages of L3 acquisition. Ringbom (2007) for example, suggests that meaning-based CLI is more prevalent in advanced learners.

One aspect that is inherently connected to proficiency is manner of acquisition. While high proficiency in adult SLA is rather common, it is unusual for adult second language learners to be native-like in all domains of the second language with pronunciation being most affected (Abrahamsson & Hyltenstam, 2009; Ortega, 2009; Rothman, Gonzáles Alonso & Puig-Mayenco, 2019). Whether this is a result of learning conditions, individual variation in aptitude, or age effects can be discussed.[10] There is a wealth of research from the past decades that has focused on the effect of type of multilingualism on cognition. Not all multilinguals are the same: there are differences depending on type of learning, context of language use,

[10] The *Critical Period Hypothesis* (Lenneberg, 1967) and other accounts of biologically derived age effects comprise a debate that is beyond the scope of this dissertation. The other two aspects of variation, namely *learning conditions*, and *individual variation*, are aspects of the language learning experience that are covered in the experiments of this dissertation.

and age effects (Luk & Bialystok, 2013). There have been suggestions that from a neurocognitive perspective, the processing of a multilingual's languages is mostly qualitatively similar. Variation in brain activation between the different languages can be mainly attributed to the use of difference cognitive resources to control inhibition rather than differences in terms of storage (Abutalebi, Cappa & Perani, 2005).

An aspect of proficiency that is important is that L2 learning outcomes are to some extent dependent on predisposition: higher *aptitude* leads to better learning results and increased proficiency leads to more available resources. Therefore the second factor of individual variation in this dissertation is aptitude, which is operationalized through its interconnected subcomponents *working memory* (WM) and *cognitive control* (CC).[11] WM is used as a cover term for a range of systems of short-term memory (Baddeley & Hitch, 1974; Baddeley, 1986). CC refers to a range of processes that facilitate selection, inhibition, and monitoring (Shallice, 1988). WM and CC have been postulated to have a major impact on the acquisition and processing of vocabulary (Ellis & Sinclair, 1996). An aspect that is highly relevant to applied linguistic theory is a proposed *supervisory attentional system* (Norman & Shallice, 1986), which is a core aspect of the aforementioned inhibition model by Green (1998). Cognitive control is also a key aspect of Baddeley's multicomponent model of WM (Baddeley & Hitch, 1974; Baddeley, 1986). The explanation for the effect of WM in explaining variance in acquisition is threefold: 1) WM is "closely related to attention" (Schmidt, 2001, p. 10) which means that individual differences in WM are related to individual differences in attention (Engle, Kane & Tuholski, 1999), 2) increased WM resources, though WMs subcomponent phonological short-term memory, explain individual variance in repetition accuracy as well as vocabulary learning outcomes (Papagano & Vallar, 1992; Service & Kohonen, 1995), and 3) WM predicts performance in listening comprehension (Harrington & Sawyer, 1992). This means that increased WM capacity allows the learners to attune to more information, to more accurately repeat verbal input, and to simultaneously process more information. Over time, these lead to better learning results.

WM and CC together are treated as *aptitude* in this thesis, which refers to a set of predispositions that predict a language learner's performance under a given set of conditions. Aptitude is hypothesized to have a larger impact on late second language learners rather than early second language learners (Ortega, 2013). All

[11] The terms *executive function* and *inhibitory control* are used in the literature for cognitive control.

four empirical studies in this dissertation focus on late second or third language learners. In this dissertation, cognitive control is approached from two different perspectives. It is treated as an independent variable by using scores from nonverbal cognitive control tasks to predict success in language acquisition. In addition, it is treated as a component of the time-course of language processing in individual trials in the experiments. An additional debate in recent decades of research in multilingualism has been on the effects of multilingualism on cognitive control (e.g., Bialystok, 2009). The assumption in this dissertation is that a hypothesized bilingual advantage is not necessarily stable, but fluid in the sense that it fluctuates depending on the demands that arise from proficiency, use patterns, and task demands (Green & Abutalebi, 2013; Green & Wei; 2014).

The third factor of individual variation in this dissertation is *psychotypology*. Language similarity as a variable in modeling multilingual language acquisition was introduced in the previous subsection. Kellerman (1983) proposed that the crucial factor for CLI in multilingual language acquisition is psychotypology, in the sense of perceived language similarity, rather than genealogical relationship (even if the two more often than not are highly correlated). Unlike genealogical relationship, psychotypology is not necessarily static. Since it corresponds to the language learner's perceived distance between the two languages, it can gradually change in the process of language acquisition. In the beginning, there is an assumption of similarity, but with increasing familiarity with the language, the learner develops a more accurate perception (Kaivapalu, 2004, Kaivapalu & Muikku-Werner, 2010; Rast, 2008; Ringbom, 2007).

As an individual variable, psychotypology can be responsible for both facilitative and non-facilitative CLI at the level of performance. Let us consider that we have a fluent speaker of Swedish and English who is trying to communicate in German and has a perception that Swedish vocabulary is more similar to German than English is. If the speaker does not know the German word for PEOPLE, 'Volk,' attempting to use the Swedish 'folk' is facilitative. This can also apply for word-internal structures. Should the learner not know the German word for LOUDSPEAKER[12] 'Lautsprecher,' but knows the translation equivalents for LOUD and SPEAKER, the learner can assume based on both Swedish and English having the same structure that this will work in German as well. On the other hand, if the speaker is trying to fill a lexical gap for BEER in German, attempting to use 'öl'

[12] In English, 'speaker' can be used to refer to a loudspeaker while both in Swedish and German neither 'talare' or 'Sprecher' carry this meaning.

from the perceived more similar Swedish will be non-facilitative as the English 'beer' would have been closer to the German 'Bier.' To make matters more dire for our example speaker, in German 'öl' refers to OIL.

In the empirical work in this thesis, psychotypology is treated as an individual variable. This means that it has to be measured, and this will be done by means of both implicit and explicit measures. For implicit measures, the participants' unconscious behavior under time-pressure is tested to estimate psychotypology through experimental means. Data for explicit measures is collected through participants' conscious estimations of similarity across languages. Explicit measures are collected for Studies 2 through 4, and implicit measures in Study 2.[13] Treating psychotypology as an individual variable includes the possibility that it is static, meaning that there is no between-participant variation or within-participant variation over time that cannot be attributed to measurement error. Measuring psychotypology, as opposed to predicting it, allows for analyzing of it as an individual variable both with respect to change over time as well as variation between individuals.

[13] Study 1 is about second language acquisition and does not include a psychotypology-component. This is not meant as a statement that psychotypology between the L1 and the L2 in second language acquisition does not matter, but rather a choice in terms of what components could be included in a limited amount of time that can be expected from the participants.

2.2 Multilingual mental lexicon

Mental lexicon usually refers to some form of a store of words. Some descriptions (e.g., Bardel, 2015; Paradis, M., 2004, 2009; Ullman, 2001, 2005) specifically refer to the mental lexicon being a store for declarative knowledge whereas others (e.g., Jarema & Libben, 2007) include procedural unconscious processes. Levelt's (1989)[14] influential model of speech processing divides lexical entries into *lemmas* and *lexemes*. These are illustrated in Figure 1 below. Lemmas are entities that contain aspects related to the word's meaning and use. They, in turn, activate *concepts*, which are building blocks of meaning. Lexemes, then, contain the form of the word (Jiang, 2002).

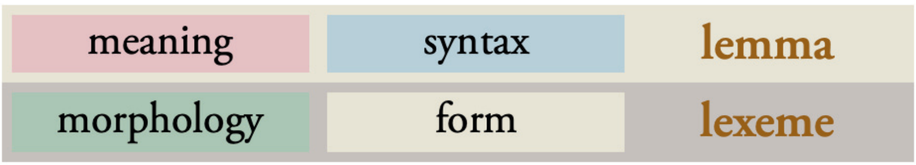


Figure 1. The lexical entry in the mental lexicon based on Levelt (1989)

It has been suggested that lexemes and lemmas are language-specific, while the conceptual system is shared across a speaker's languages (Gabrys-Barker, 2005; Jarvis, 1998, 2000; Kellerman, 1995; Slobin, 1993). These three levels: concept, lemma, and lexeme form a *lexical triad*. In instructed post-L1 language learning there is often an assumption that vocabulary learning entails learning new labels for existing lexical knowledge, something that Falk (2010) refers to as *relexification process* in the case of related languages. Ample work has been done (see list in section 1.1) on CLI that presents itself at the level of the lexeme, i.e., at the surface level, irrespective of its origin. Empirical work in this dissertation focuses on the interface between what in Levelt's classification would be the lemma and the conceptual store.[15] This type of CLI is, by nature, more subtle than that which takes place at the level of the lexeme or at the interface between the lemma and the

[14] An adaptation to bilinguals has been proposed by De Bot (1992).

[15] *Concepts* and *conceptual* are used in this dissertation instead of the related, and broader term, *semantic* which is often used in linguistics but which in psychology is used to refer to all types of factual knowledge.

lexeme in the speaker's languages in choosing an incorrect surface form due to a mismatch between the speaker's languages.

Issues caused by a mismatch between meaning across languages have been noted already in early models of bilingual word recognition (Weinreich, 1953). Developmental models of the MML typically start from the assumption that a language learner is heavily dependent on L1 lemmas at early stages of word learning (Jiang, 2002; Kroll & Stewart, 1994; Pavlenko, 2009; Potter, So, Eckardt, & Feldman, 1984). A speaker can have different compensatory processes to deal with these mismatches which, based on the aforementioned assumption, would surface in the target language in the early stages and for all the speaker's languages at a more advanced stage in the case of a converged system. One possible compensatory process is subconscious mitigation of unavoidable CLI which is caused by co-activation of the speaker's languages. This would take place by inhibition of activation levels of task-irrelevant components. On the other hand, the speaker – should the time-course so permit – can consciously apply declarative information about the differences between word meanings in different languages to adjust the use of a particular lexical item. These may hinder qualitatively similar errors in language production, but represent different types of language processing with the former being more automatized. Mitigation would present itself primarily in on-line processing data while conscious application can also be observed in slips of tongue and other forms of non-target-like use of language in production data.

It should be clear that what happens in the mental lexicon is much more complex than a simple activation of meaning based on initial conceptual construction. In this thesis, *multilingual mental lexicon* (MML) is used as a metaphor to refer to the complex range of interconnected memory systems, conceptual knowledge, heuristics, as well as conscious- and unconscious cognitive processes of activation and retrieval that together form our capacity to use lexical items across all our languages. The following sections will cover the building blocks of word meanings starting with *categories* and *prototypes,* which are building blocks of *concepts*, and then concepts, nouns, and finally semantic relationships between nouns. Relationships are important since when the speaker's languages do not match in word meaning, it is often the case that a single lexical item corresponds to two or more lexical items in another language, which then, in turn, have a semantic relationship within that language.

2.2.1 Building blocks of nominal meaning

The empirical studies in this dissertation look at acquisition and processing of nouns in multilingual speakers. A primary focus is on change in unconscious lexical activity over the course of acquisition. For someone wanting to research subtle changes in a language user's vocabulary, the inclusion of changes that might not represent themselves at the surface level, or be known to the language speaker consciously, is of utmost importance. Since language, in its essence, is a product of human cognition – as well as acts as an instrument aiding it – it mirrors the structure and functioning of general cognitive abilities (Taylor, 2003). In the previous section, it was suggested that mental entries for lexical items are divided into a lexeme, a language-specific lemma, and language-independent conceptual system. Most developmental models of MML start from such assumption. In terms of proposals of the connection between the lexeme, the lemma, and the conceptual system, we have three distinct proposals. The most modular accounts differentiate between the semantic meaning of a particular word and general conceptual knowledge (e.g., Bierwisch & Schreuder, 1992; also see Pavlenko, 2009 below). This is referred to as *two-level semantics*. Levelt's (1989) model is a so called *one-level semantics* model in that it assumes semantic and conceptual knowledge to be the same. Levelt's model still departs from autonomous and specialized modules (albeit referred to as *components*). The least modular account is *cognitive linguistics* (e.g., Paradis, C., 2012) which suggests that there is a direct mapping of lexical forms to conceptual structures. This account highlights the constant state of change and the fluidity of word meaning since word meanings are construed online based on the context of their occurrence in each individual use situation. At the end of the day, whether there is a separation of linguistic meaning and non-linguistic meaning is an empirical question, one which to a large extent is beyond the scope of this thesis. However, given the focus on change of meaning representations in multilingual speakers, it is important to acknowledge that such a debate exists.

The two most basic building blocks of conceptual knowledge are proposed to be *categories* and *prototypes*. These form the basis of the conceptual system. Situations that we encounter in our surroundings differ vastly on a day-to-day basis. In order to cope with this, humans categorize entities. The prime purpose of this categorization is to reduce the complexity of our environment. For linguistics, *categorization* is interesting for two reasons: 1) categorization involves naming (onomasiology), and 2) language is an object of categorization (Taylor, 2003). Rosch

(1978) has suggested that the "task of cognitive systems is to provide maximum information with the least cognitive effort" (p. 28). One proposal is that categories are exemplified by a *prototype*. Consider the word 'cup.' You will likely not have any difficulties visualizing a typical cup. The prototype you envision serves as a reference point when you categorize instances that are not equally clear. This, however, does not mean that prototypes of categories are universal. Rather, sometimes the prototypes have details that are culturally bound: for example, different animals would be prototypical in different regions due to their varying distribution. Furthermore, sometimes prototypes are contextually bound. Rosch (1973) found that the degree of category membership affects decision making time and priming. It takes less time to decide that a robin is a bird than that a duck is a bird.

Now that we have considered the different building blocks and processes of the conceptual system, we can move to the next building block of the mental lexicon: *concepts*. Concepts are non-linguistic representations, or mental entities, "that form systems of areas of human experience" (Paradis, C., 2015). Earlier in this dissertation it was mentioned that no sensitive periods have been proposed for lexical knowledge.[16] If word meanings for a monolingual are not stable but are seen to depend on their use, then it is plausible that word meanings are susceptible to change in meaning also due to the multilingual experience. Langacker (1987) takes the stance that there are no speakers that share the same linguistic system with other speakers. This is due to the dynamic nature of meaning where each usage event affects the structure of the invoked meanings. This aligns well with the aforementioned proposal of *retrieval-induced forgetting* (see section 2.1). Say that our example speaker uses Finnish, Swedish, and English on a daily basis. Every time the speaker uses the Finnish word 'sitruuna,' that invokes the category CITRUS FRUITS as well as the translation equivalents: Swedish 'citron' and English 'lemon.' This coactivation affects the structures of these invoked categories. Furthermore, since a part of these representations are shared with, say, oranges, the invoked networks are also affected. Since activation of the related traces requires for them to be suppressed, over time, these traces are affected. Langacker (1987) suggests that a similar process is the driving force of the initial acquisition of lexical

[16] In many ways, for vocabulary, the situation is the opposite in that instead of sensitive periods in childhood when vocabulary learning is postulated to take place, an ever-growing *world knowledge* (crystallized intelligence) associated with increased age and life experience allows us to acquire the meaning part of new words more easily.

networks and any nodes in those networks are continuously *entrenched* by subsequent activations of the network. Hence, changes in the experiences and communicative needs alter the configurations of these networks even in mature speakers. Entrenchment refers to the routinization and automation caused by activation of "cognitive occurrences of any degree of complexity" (p. 100). Each use of a structure "has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact" (p. 59). From this perspective, frequency of activation would be the same as the level of entrenchment (Schmid, 2007). An alternative take – which is related to the aforementioned RIF – is that frequency of activation of a particular cognitive unit in comparison to the frequency of activation of competing cognitive units determines the level of entrenchment (Geeraerts, Bakema & Grondelaers, 1994).

Taylor (2003) suggests that differences in representations between speakers usually go unnoticed in communication. It is plausible, though, that these do occasionally lead to barriers in communication or misunderstandings. For us to be able to communicate effectively, the meanings of the words must be shared by the speakers of the language. The question is then, if everyone has their own individual concepts, how is it that we are able to communicate (c.f., Fodor, 1977)? There are several caveats with this line of thinking. First, people attach word forms to existing concepts. If one's communication is unsuccessful, the speaker is forced to revise the associated conceptual structure through a social process of convergence between speakers leading to highly *conventionalized* meanings.[17] Secondly, this process works both ways, whereby the society is affected by its individual speakers, and the speakers are affected by their interlocutors (Clark, 1996). In the end, this results in the speakers – within a given language community – having rather comparable concepts. As such, the assumption that concepts are neither shared nor stable does not hinder effective communication (Murphy, 2002). One category of words that have relatively easily described conventionalized meaning is nouns, particularly concrete nouns. On the other hand, this is probably the type of word that most easily lends itself to be taught to second language learners without using linguistic means – for example, using pictures – and where the assumption is that there is a one-to-one match between languages.

---

[17] *Conventionalized* here refers to a social process rather than a core meaning which is a "decontextualized and presumably invariant concept associated with a word" (Anderson & Nagy, 1989).

Nouns differ from other lexical items in that they often refer to highly coherent bits of the world. Particularly concrete nouns refer to naturally individuated referents (Gentner, 2006). More specifically, based on Talmy's (1975) findings of verb variability across languages, Gentner (2006) proposed that verb meanings are more variable across languages than noun meanings and that verbs, in general, are more variable in their semantics than nouns, even within languages. This, then, is one potential explanation to why children learn nouns before verbs. This hypothesis is called *relational relativity* (Gentner, 2006). The hypothesis makes two predictions: 1) that there is a universal early advantage in acquisition for nouns, and 2) that possessing these early nouns helps children with the learning of less transparent relational terms such as verbs and prepositions (p. 546). Some nouns are ambiguous in the sense that the mapping is something else than one word, one concept (Murphy, 2002). The question is, then, how do we represent such pairings? What is the relationship between the forms that both attach to one concept? How about two distinct conceptual representations that map into a single form? These relationships can be formulated and categorized through different *sense relations*.

Sense relations between lexical items within one language can be divided into three basic categories: a) same form but different meanings e.g., *homonyms* and *polysemes*, b) similar meanings but different forms e.g., *synonyms* and *hyponyms*, and c) different forms and different meanings but relations through opposition (Paradis, C., 2012). This dissertation focuses particularly on the operation of polysemes in multilingual language contexts. Polysemy is typically seen as a word form that takes on two or more *related* meanings. To give an example, the English word 'head' can be used for TOP OF THE HUMAN BODY, TOP OF AN ORGANIZATION, or TOP OF A BEER. The common denominator here is the underlying conceptual meaning of AT THE TOP. In Finnish, on the other hand, these three meanings do not map onto the same form. Finnish uses 'pää,' 'johto,' and 'vaahto,' respectively. We can divide the first subtype, i.e., lexical items with the same form but different meanings, into those with an *arbitrary* or *motivated* relation (Paradis, C., 2012). Word pairs that share the same form, but have an arbitrary connection with respect to meaning, are called *homonyms*. By arbitrary, it is meant that they just happen to share the same form. One example of such a word is the English 'bank,' which can refer both to i) a MONETARY ORGANIZATION and ii) SLOPED LAND. It would be hard to conceive that both meanings share a single concept. However, at that point we would need to clearly define *single concept.* A highly interconnected system of knowledge does not exactly operate in that way (Murphy, 2002). Even homonymous items seem to elicit higher conceptual similarity than otherwise comparable form pairings (Suhonen, 2015). Reduction of the conceptual representation of items to not include their connections (be it associative or those at the form level) would not exactly represent the way conceptual activation operates. The more plausible take is to consider that a lexical item elicits (through activation and inhibition) a coherent part of a concept and its connections that suits the context it is used in (Murphy, 2002). Motivated pairings, then, on the other hand, are related conceptual associations and are often referred to as *polysemes* (Paradis, C., 2012). They are essentially unambiguous words with complex meanings.

Most content words in English have multiple interpretations (Murphy, 2002). Tuggy (1993) questions whether it is possible to draw absolute boundaries between ambiguity, polysemy, and vagueness. The presumption is that ambiguity and vagueness occupy the two extremes of the continuum and polysemy lies somewhere in the middle. In this dissertation, the homonymy-polysemy categorization is assumed to be continuous variation of motivatedness, with no clear cut-off point, rather than a binary distinction.

2.2.2 The developing multilingual mental lexicon

Ausubel (1964) suggested that it should be easier for adults to learn a new language since they "need not acquire thousands of new concepts but merely new verbal symbols representing those concepts" (p. 421). In many ways, this is how we learn new languages in our daily lives: by learning new forms and attaching them to existing form-meaning mappings in languages already in place. The caveat with this line of thinking is that it assumes one-to-one matching with lexical-conceptual relationships across languages. This is far from the case, though, with estimates of words with more than one translation equivalent in another language ranging from 25% (Tokowicz, Kroll, de Groot & van Hell, 2002) to 69% (Prior, MacWhinney & Kroll, 2007). Perhaps the most cited example, presented in (1) below, of CLI in production data comes from Ringbom (1987) where a speaker of Finnish used the word 'language' to refer to TONGUE in

(1) "He bit himself in the language" (p. 116)

Both 'language' and 'tongue' share the same translation equivalent 'kieli' in Finnish. The case is particularly difficult since in English 'tongue' can be used to refer to both LANGUAGE and TONGUE, but 'language' can only refer to LANGUAGE.

In classroom-based instruction students are taught words and their associated meanings at the level of explicit knowledge. This does not necessarily lead to automatic processing (Pavlenko, 2009) even if deliberately learned words can become represented implicitly (Elgort, 2011). There has often been an assumption that when learning a new language as an adult, one learns vocabulary by mapping existing concepts to L2 word forms via L1 word forms. There is, though, at least a theoretical assumption that in the case of highly fluent individuals, L2 acquisition can change the L1 (Pavlenko, 2009). If we want to measure this, we need to tap into conceptual restructuring since, as pointed out earlier, CLI does not only present itself in overt errors (Jiang, 2002). Some researchers, like Jiang, are pessimistic with respect to the prospects of late L2 learners acquiring native-like form-meaning mappings. Jiang (2000; 2002) proposes that it is likely that most vocabulary items in the L2 fossilize to represent information that is primarily derived from the L1 and that a learner's knowledge about cross-language lexical ambiguity is typically explicit by nature.

Jiang (2000) has proposed a three-stage system for vocabulary acquisition in second language learning, which is outlined in Figure 2 below. This developmental model has continuous progression between the stages, i.e., there are no distinct boundaries.

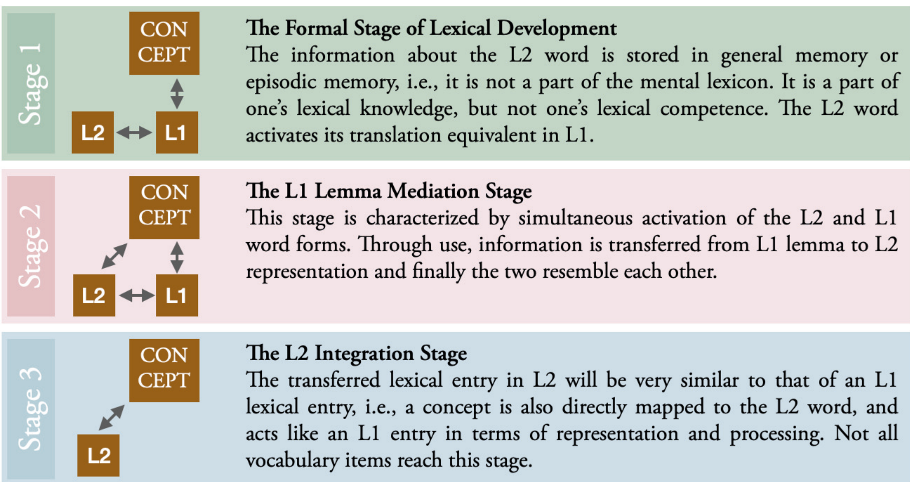


Figure 2. A three-stage system for vocabulary acquisition in SLA (Jiang, 2000, pp. 51-54; adapted from Suhonen, 2015)

It is also possible that a learner does not go through all the stages: particularly the third stage is beyond most learners and most vocabulary items. Jiang's third stage is a highly hypothetical end-state by its nature in a highly connected multilingual mental lexicon. In Suhonen (2015), this bilingual model was put to test in multilingual participants. The mental representations of Finland-Swedish bilinguals' English words were not monolingual-like in comparison with Swedish, Finnish, or English, but rather seemed to be a form of a *hybrid* that was dependent not on proficiency but factors like activation/dominance, mother's language skills, and psychotypology. The results could be better explained by a model, such as the Revised Hierarchical Model below, that assume cross-language activation even at later stages of language acquisition. However, since no time-course data was collected, it is impossible to evaluate to what extent these results relate to representation, metalinguistic knowledge, and cognitive control. In the empirical studies in this dissertation, it should be possible to differentiate between these.

Another influential model of bilingual lexical development is the proficiency-driven Revised Hierarchical Model (RHM) by Kroll and Stewart (1994). The model builds upon the Hierarchical Model by Potter, So, Eckardt, & Feldman (1984). Support for this model comes from asymmetry in processing times when comparing translation from the L1 to the L2, and from the L2 to the L1.

The RHM emphasizes shared conceptual store. It describes the development of processing of individual words over the course of acquisition. The main developmental hypotheses are the following: 1) In the early stages, lexical items in the second language are connected to their translation equivalents in the first language. Conceptual information is activated via a L1-link. 2) In the course of increasing proficiency, direct links between the second language lexical item and the conceptual store develop gradually (Kroll & Stewart, 1994). The importance of this model to the empirical studies in this dissertation is that it proposes testable hypotheses with respect to the operation of languages beyond the first language in a developing MML, namely that there is a single conceptual system and thus changes in that system affect a multilingual's all languages. However, it is unclear how a third language would operate in this model. The model is presented in Figure 3 below.

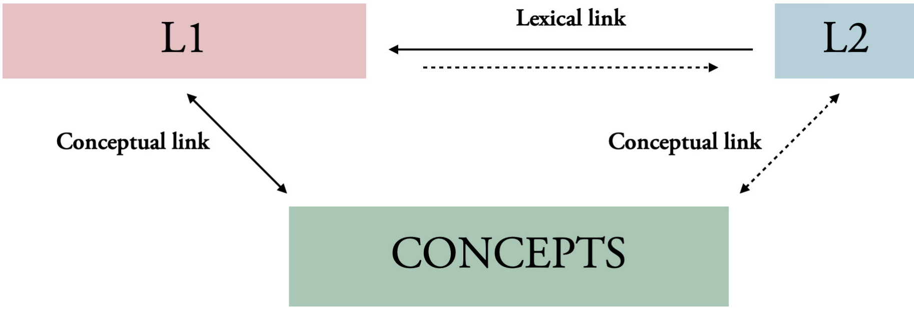


Figure 3. The Revised Hierarchical Model

The arrow from L2 to L1 represents a lexical link and is caused by a translation approach. The dotted arrow below represents a lexical link from L1 to L2. At low L2 proficiency, this connection is weak. Increasing proficiency in L2 coincides with stronger activation from L1 to L2. The arrows between the L1 and the conceptual level, and the L2 and the conceptual level represent links between lexical and conceptual information. At lower L2 proficiency levels, access from L2 is via L1 by default. When proficiency in the L2 increases, the direct L2 to conceptual storage-link becomes stronger (Kroll & Stewart, 1994, p. 150).

Like Jiang's three-stage system, the RHM assumes conceptual equivalence. Some models of multilingual lexical processing and development suggest that there is some level of language separation in the conceptual system. One proposal is that the degree of interconnectedness in the lexical representations across languages is dependent on the type of lexical item. The Distributed Feature Model (van Hell & de Groot, 1998) proposes that concrete words have a more shared representation across languages than abstract words do.

The Distributed Feature Model is presented in Figure 4 below. The colored circles refer to shared features and the white circles to separate features. Support for this model comes from bilinguals' translating of concrete nouns being faster than that of abstract nouns (Van Hell & De Groot, 1998). To maximize potential overlap as well as to control representations, in the present thesis, stimuli in all four experiments is based on concrete nouns.

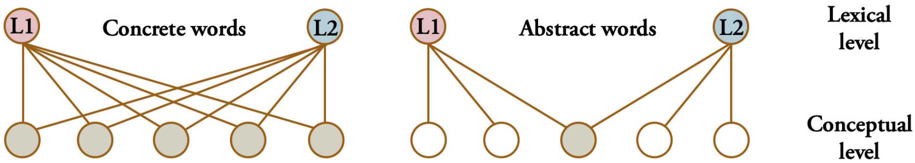


Figure 4. The Distributed Feature Model (Van Hell & De Groot, 1998, p. 205)

The aforementioned proposal of conceptual non-equivalence has also been brought up within a frame of a developmental model. Pavlenko (2009) has proposed a modification of the RHM – Modified Hierarchical Model (MHM) – that emphasizes variation in conceptual equivalence. In this model, conceptual aspects of languages are not fully shared but instead are formed based on separate and shared categories. Learning in this model is a gradual, implicit process, and includes conceptual restructuring. Pavlenko refers to restructuring as "changes in speaker's linguistic categories, seen as a subset of cognitive categories" (2011, p. 246). The MHM allows for a possibility for independent conceptual development in all of a speaker's languages, and as such aligns with the proposed end-state in Jiang's aforementioned model. Pavlenko's MHM is presented in Figure 5 below. The arrow from the L1 to the L2 represents a lexical link. This is due to a presumed translation approach in words in the L2. The dotted arrow in the opposite direction represents a lexical link from the L1 to the L2 and this link, particularly at low proficiencies, is suggested to be weak(er). Increasing proficiency in the L2, then, also means that the link becomes stronger. The dotted arrows represent possibility between direct connections. Conceptual development, which affects both the L1 and the L2, takes place at the conceptual level (Pavlenko, 2009, p. 147).

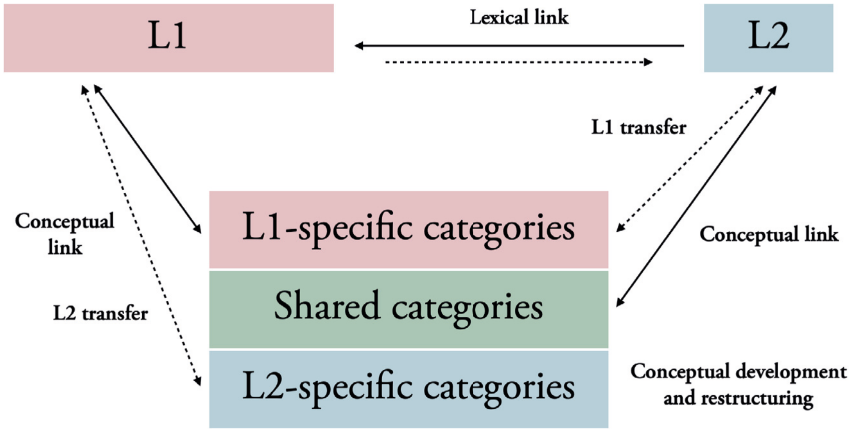


Figure 5. The Modified Hierarchical Model

The predictions of the aforementioned developmental models also entail that language learners have variation in their *interlanguage*. Selinker (1972) proposed, in line with Weinreich (1953), that language learners' developing language is systematic even if it is not target-like. This language is referred to as interlanguage. Here, then, the important theoretical and empirical take in research on L3 acquisition of sequential bilinguals is that one should not compare features of a given L2 system on the basis of assumed monolingual native speaker-norm without establishing properties of the individual's interlanguage (Rothman, 2015, p. 188). Furthermore, in case of multilingual language learners, Rothman and Cabrelli Amaro (2010) highlight the need to verify that the features of a given language have not undergone CLI or attrition before assuming that they can be transferred to the L3.

The lexicon has been proposed to be the first affected area in language attrition (Ecke, 2004; Schmid & Köpke, 2008). Ecke (2015) proposes that CLI plays an important role in this and that multilingual competence is in constant state of change. Schmid and Köpke (2017) argue that attrition covers both 1) the L1 becoming less accessible and 2) being modified due to the multilingual experience. It should be possible to differentiate between these effects from a perspective of causality. If a native speaker of Finnish has difficulty retrieving the form 'tuoli' CHAIR due to not having used that lexical item for a long time, this is caused by the L1 becoming less accessible. However, if the same speaker starts using the form 'sääri' LOWER PART OF THE LEG to refer generically to the whole LEG as a result of conceptual restructuring then this effect is a result of CLI, i.e., it is not caused by the lack of co-activation of the Finnish form but a result of repetitive co-activation of the non-matching conceptual representations. Retrieval difficulties can also be caused by 1) unintended interfering intrusion of another language form, or 2) temporarily impaired or incomplete representations (Ecke, 2015). In real-life situations it is likely that a multilingual speaker whose primary language use has shifted from the mother tongue to another language is subject to all four aforementioned processes that cause occasional temporary retrieval difficulties.

The MML in speakers with more than two languages is different from monolinguals and bilinguals in that the additional languages increase the number of possible sources and directions of CLI. This is due to the more complex configuration of the lexical networks (Ecke, 2015). Hall (2002) has proposed that one of the default cognitive processes of lexical development in the MML is *parasitism,* which refers to a strategy of searching, detecting, and using similarity to "integrate novel word structures into a network of stable representations and access routes" (Ecke, 2015, p. 149).

The Parasitic Model, henceforth PM, (Hall & Ecke, 2003) suggests that CLI takes place at the level of *form*, *frame*, and *meaning* (c.f., lexeme, lemma, and concept). Furthermore, there are default processes and stages in the process that relate to aspects of the lexical item's developmental trajectory as it integrates into existing lexical networks. From this aspect, PM is similar to the aforementioned developmental models by Jiang (2000), Kroll and Stewart (1994), Pavlenko (2009), and Weinreich (1953). PM has three primary stages: 1) establishing a form representation, 2) building connections to frame and concept representations, and 3) strengthening and automatization of representations and access routes. Each of these has a series of secondary stages. These are presented in Table 3 below.

The most important addition by this model is that it highlights that this process takes place for each lexical item individually (item-by-item as opposed to wholesale transfer). Each lexical item is subject to its own developmental trajectory. These trajectories are not linear by nature but include lapses and restructuring. Furthermore, creation, revision, and abandonment of connections takes place at all three levels of the lexical triad (form, frame, and meaning). Thus, different lexical items can be at different stages in the emerging MML simultaneously and this process is characterized by occasional issues in production (Ecke, 2015). The model does not make any explicit predictions to the effect of later acquired languages on earlier acquired languages. From the perspective that multilingual language learners learn new items in the L1s throughout the lifespan this is unfortunate.

Table 3

Stages in the Parasitic Model of Lexical Development (Hall & Ecke, 2003)

	Establishing form		Building connections		Automatization
A1	The L3 word form is registered in STM and the closest matches in L3, L2, or L1 are activated based on salient form attributes.	B1	The frame of the form-related host is adopted for deployment of the L3 form. It is retained while contextual cues confirm the inference, as it is used as a link to the corresponding conceptual representation.	C1	Initially established connections with other L1, L2, or L3 representations are revised, bypassed or severed, to establish a more autonomous triad responding to new cues in input.[18]
A2	The L3 form is connected to a host representation[19] and is established in LTM in distributed fashion.[20]				
A3	Differences between L3 form and host representation are detected, new patterns are rehearsed and the representation is revised with respect to the attributes that distinguish it from the host and/or other consolidated neighbors.[21]	B2	If subsequent context contradicts information in the frame and conceptual representation inferred from the form-related host, another perceived conceptual equivalent from L1 or L2 is activated.	C2	Autonomous connections between L3 form, mediating frame, and concept are strengthened and the representations themselves refined, with increased frequency of exposure and use.
A4	If no matching representation is activated sufficiently, the L3 form is connected to the frame of the nearest conceptual equivalent.	B3	If no translation equivalent can be identified, a provisional frame[22] is constructed and connected directly to a conceptual representation.	C3	Access routes between elements of the L3 triad are automatized.

Note. STM refers to short-term memory and LTM to long-term memory. The three bold levels at the top refer to the three stages of development (adapted from Ecke, 2015).

[18] This is not always achieved, leading to *fossilization* (cf. Jiang, 2000).

[19] This is normally the most highly activated related L3, L2, or L1 form, where some threshold level of similarity between the items is met.

[20] This refers to the activation of the same nodes in the network as the host form.

[21] Not always achieved, leading to fossilization of the interlanguage configuration (cf. Jiang, 2000).

[22] Based on a variety of distributional and morphological cues.

In PM, CLI is modulated by a set of factors: *learner factors, learning factors, language factors, event factors, and word factors.* The components of these factors are presented in Table 4 below and are representative of many of the factors presented in the previous sections in this thesis.

Table 4

Modulating Factors in the Parasitic Model of Lexical Development (Hall & Ecke, 2003)

Type of factor	Factor
1. Learner	Psychotypology, metalinguistic awareness
2. Learning	L2 status, proficiency in each language, acquisition order
3. Language	Typological distance, degree of contact between languages
4. Event	Language mode, task, style, interlocutor
5. Word	Degree of similarity (form, frame, concept), abstractness vs. concreteness, frequency, frequency of competitors

The first, *learner factors* refers to those modulating factors that have within-subject and between-subject variation in language learners. Ecke (2015) does not mention aptitude under this area but that would fall under learner variation. *Learning factors* are related to learner factors but the focus of those is more on aspects of the learning experience: for example, which language one acquired first, whether it was a second or a third language, and how proficient one is in each language. These have significant overlap, but it is not uncommon to be more proficient in the L3 than in the L2. *Language factors* have no within-subject or within-participant variation. These are factors that relate to linguistic (genealogical relationship) or societal (degree of contact between languages). To give an example here, Finnish and Swedish are languages that are geneologically very distant but that have a lot of societal contact since both are widely spoken in both countries. This, then would result in a fair amount of lexical overlap despite genealogical distance. *Event factors* refer to aspects that vary in a particular language use situation. If both interlocutors are fluent in several languages, both languages are more likely to be active. On the other hand, different words are used in different registers. Finally, *word factors* relate to variation in the lexical items. Particular importance for this thesis, is the aspect of degree of similarity at the form, frame, and conceptual level.

This is another strength of the PM, as it makes predictions to why CLI occasionally causes surface level effects and why it occasionally doesn't even if the underlying representation would not have changed. This is also the primary relevance of the PM for this dissertation as this dissertation looks at all potential six directions of CLI in the MML in third language acquisition. The modulating aspects can be applied to all the six directions. The model also makes predictions with respect to developmental stages for individual items in the MML. This is also relevant to this thesis, but the model only makes predictions about forward CLI in the L3. For the other directions of interest (lateral, reverse), there are no models that make specific predictions at this level.

2.3 Previous research

2.3.1 Resolving translation ambiguity

Translation ambiguity refers to one-to-many mappings between words in a learner's languages (Eddington & Tokowicz, 2013). An example of translation ambiguity is Finnish 'maali,' which refers to both PAINT and GOAL in English. Other terminology has also been used in previous research, such as *cross-language differences in lexical-conceptual relationships* (Elston-Güttler, 2000), *meaning extensions* (Odlin, 2008), and *different-translation pairs* (Jiang, 2002). Elston-Güttler (2000) points out that the lack of one-to-one mapping in lexical-conceptual representations is often not well established in models of multilingual lexical processing. However, translation ambiguity has been found to be highly common. For example, Prior, Kroll, & MacWhinney (2013) found that 50% of Spanish words and 60% of English words have some level of translation ambiguity across the two languages.

Given that translation ambiguity is so common, how multilinguals resolve it in language processing is an important question for research, particularly when modeling CLI in the MML. All four studies in this dissertation look specifically at learning and processing of words with translation ambiguity. Previous research on four relevant aspects is presented below: *resolving*, *learning*, and *processing* words with translation ambiguity as well as proposed factors *modifying* their processing.

With respect to *resolving* translation ambiguity, The Revised Hierarchical Model, RHM, was presented in Section 2.2.2. The RHM suggests that – at low proficiency in the L2 – links from the L2 words to the L1 translations are stronger than vice versa. Also, the assumption is that links between concepts and L1 words are stronger than between concepts and L2 words. This leads to translation from L2 to L1 taking place via lexical associations while translation from L1 to L2 takes place via concepts. The RHM does not, however, take translation ambiguity into account. Eddington & Tokowicz (2013) proposed a modified version of the RHM, the Revised Hierarchical Model of Translation Ambiguity (RHM-TA) that makes developmental predictions for items with either form- or meaning based translation ambiguity. An example of a word with a *form-ambiguous translation* is Finnish 'puku,' which translates to highly related SUIT and COSTUME in English. An example of a word with *meaning-ambiguous translation* would be the aforementioned Finnish 'maali,' which translates to unrelated PAINT and GOAL in English thus constituting a case of homonymy. The model is presented in Figure 6 below. In the model, intralingual status of the items (gradable from synonymy to homonymy) defines the type of interlingual translation links.

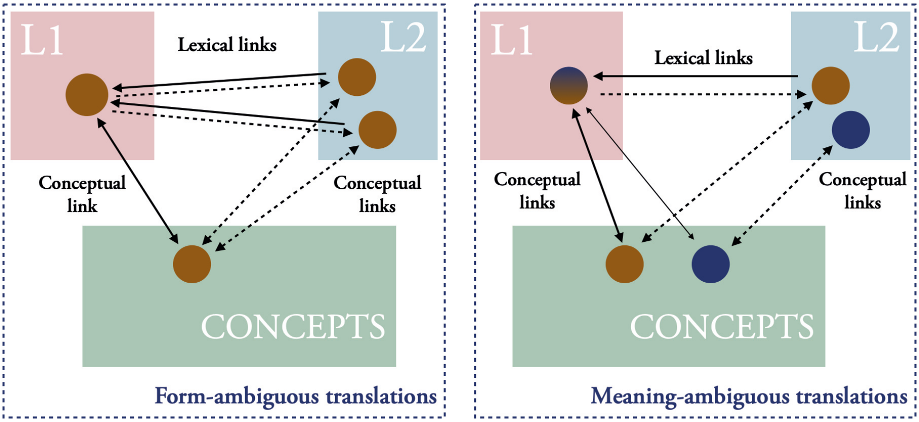


Figure 6. The Revised Hierarchical Model of Translation Ambiguity

The difference between form- and meaning ambiguous translation ambiguity has to do with the relationship of the two translations for the L1 word. In the form-ambiguous type (left panel), the two translations in the L2 are highly related representing near synonymy. In the meaning-ambiguous type (right panel), the two translations in the L2 are not related and are thus represented by separate concepts even in the L1 (adapted from Eddington & Tokowicz, 2013).

In the literature, there have been both proposals for translation links at the level of the lemma and the lexeme (see Figure 1 in Section 2.2), as well as hypotheses with respect to at which of the two levels inhibition takes place. Starting with the former, in an influential early study of word processing, Potter, So, von Eckhardt, and Feldman (1984) compared two hypotheses of bilingual lexical and conceptual representation: 1) that associations are established between words in the two languages, and 2) that associations are formed via concept mediation. Earlier hypotheses had primarily been in favor of the first explanation, while the results of Potter and colleagues supported the second hypothesis. Elston-Güttler, Paulmann, and Kotz (2005) tested processing of translations of L1 homonyms such as 'Kiefer' in German that refers to both PINE and JAW using Event Related Potentials (ERPs). In a low-proficiency group, the results indicated lexical-level translation, and in a high-proficiency group, the results indicated independent form-meaning mappings. The results align with the predictions of the development proposed in the RHM-TA. This also means that both aforementioned hypotheses were confirmed, with low-proficiency participants operating in accordance with the first hypothesis, and high-proficiency participants with the second. In relation to inhibition, Lee and Williams (2001) as well as Meuter & Allport (1999) tested cross-language competition effects (see 2.3.2). They found that while there were cross-language competition effects, there were also strong effects of inhibition of the L1. Lee and Williams (2000) suggest that this inhibition might take place at the level of the lemma.

It is possible that language learners have to resort to conscious rules to resolve translation ambiguity. As pointed out earlier, a lion's share of research on CLI in the MML has focused on production errors. Jiang (2002) has suggested that this is problematic and hypothesized that the LTM might contain explicit knowledge, 'instructions,' about how to use particular words where there is a lack of conceptual non-equivalence. These would be, presumably, acquired through explicit instruction. Jiang (2000) specifically suggests that these explicit rules are not a part of the mental lexicon. Jiang's proposal, then, means that given adequate time and cognitive resources in a particular use situation, a learner can correct their use of a particular L2 word. To give an example, Swedish (neither do Finnish nor German) does not lexicalize any difference between TO LEND and TO BORROW. Both directions of the act are referred to by the form 'att låna' in Swedish. Assuming that a L1 Swedish learner of L2 English is dependent on their L1 conceptual representation, in order to perform accurately in English, they then employ a conscious rule to adjust their use of the two items in the L2 English. This type of conscious adjustment, given adequate time and cognitive resources, will be referred to as *introspection* in this thesis.

The second relevant area of research is *learning* words with translation ambiguity. In general, bilinguals have been found to be better at novel word learning than monolinguals (Kaushanskaya & Marian, 2009; Papagno & Vallar, 1995; Van Hell & Mahn, 1997). Kaushanskaya & Rechtzigel (2012) found that this effect seems to be, at least partially, explained by greater sensitivity to semantic information during learning. Bogulski, Bice & Kroll (2018) tested whether English-Spanish bilinguals are advantaged in learning new words in L3 Dutch compared to monolinguals who spoke only English. They found that the bilinguals were slower in the learning tasks, while they outperformed the monolinguals in learning outcomes, suggesting that the extended time (or difficulty) in the learning resulted in better learning outcomes. Both groups learned the additional language using the L1 (i.e., using translation equivalents). Later, when tested whether the effect was the same for the bilinguals when the learning took place in the L2, the effect of longer time on task in learning and better outcomes disappeared.

It seems that learning items with translation ambiguity is more difficult for second language learners than items without translation ambiguity (Degani & Tokowicz, 2010b; Eddington & Tokowicz, 2013). To test whether this effect is modulated by *learning type* variables, Degani, Tseng, and Tokowicz (2014) taught Dutch words with and without translation ambiguity in English to native English speakers with the aim of testing whether type of instruction can alleviate these effects. Translation ambiguous words had two word forms in the target language and the translation unambiguous ones one. The training sessions were manipulated so that it was possible to compare whether close proximity (operationalized by the amount of other items presented in between) in presentation of the two target language forms (as opposed to spacing) would produce better learning results. A particularly relevant finding for Studies 1 and 2 (Chapters 3 & 4) in the present dissertation is that Degani, Tseng, and Tokowicz (2014) found primarily effects of translation ambiguity in *accuracy* – but not in *processing* – when testing learning outcomes. Furthermore, presenting ambiguous items in close proximity led to better learning outcomes. Also, translation ambiguous items were recalled better in two delayed post-tests administered one and three weeks after training.

In addition to being more difficult in learning, words with translation ambiguity have also been found to be more costly in *processing* (Degani & Tokowicz, 2010b; Tokowicz & Kroll, 2007). Eddington and Tokowicz (2013) offer two explanations for increased processing cost associated with translation ambiguity: 1) building on the *fan effect* (Anderson, 1974), the association strength between words with multiple meanings is smaller than for words with only a single associate. Thus, in the case of translation ambiguity, given an increase in the number of translations, the association between a word and its translation equivalents is smaller. In terms of a second explanation, 2) *Interactive activation* accounts (e.g., McClelland & Rumelhart, 1981) would suggest that competition takes place between translations that connect to one word. Selection, then, would require inhibition of the task-irrelevant competing stimuli. In line with the latter explanation, based on the predictions of Kroll & Stewart's (1994) RHM, the activation of items in the shared conceptual system should activate the lexical representations in the speaker's other languages as well.

Elston-Güttler (2000) found that the processing effects of non-equivalence in lexical conceptual relationships across a speaker's languages depend on 1) the type of lexical relationship the words have, and 2) the context the lexical processing takes place. Translated meanings of interlingual homographs only primed in the L2 when presented in isolation, i.e., not in context. Translated meanings of items with translation ambiguity with a homonymous relationship in the L2 primed in the L2 when presented both in and outside of context. For items with translation ambiguity that were polysemous in the L2, there were both more overt errors as well as longer response times at the group level when tested in the L2 in a sentence anomaly judgement task. That the aforementioned effects were modulated by the degree of relatedness in the L2 support the RHM-TA.

The presented results on processing, when taken together, suggest that L1 form-meaning mappings do indeed have an effect on the processing of L2 words. However, whether bilinguals utilize primarily L1 conceptual representations in combination with explicit knowledge to resolve translation ambiguity (in line with Jiang, 2002), or whether the processing effects are caused by co-activation of the L1 form-meaning mappings even when direct L2 form to concept links have been formed – in line with Elston-Güttler & Williams (2008) – is difficult to differentiate based on time measurements alone. Presumably, in the case of the former, the observed increase in processing cost would be larger (with explicit rule application taking more time than what lexical competition effects would add) than in the latter case. In both cases we expect to see a relationship between response times and accuracy. In the former case, time spent on applying explicit knowledge leads to more target-like outcomes, whereas in the latter case co-activation of lexical and conceptual information in the other language could potentially result in less-target like outcomes.

The effects of translation ambiguity in acquisition, representation, and processing are likely *modulated* by both learner and item related factors. Laxen and Lavaur (2010) examined the effect of translation ambiguity – specifically the effect of modulating factors *number of translations*, *dominance of translations*, and *semantic relatedness* – in French-English bilinguals. In a translation recognition task, the participants had to decide whether two items, one from each language, were translations or not. All of the three aforementioned factors showed an effect in response times in the task. In a productive translation task, Prior, Kroll, and MacWhinney (2013) found that language users tended to choose *cognates*, if present, over other potential translation equivalents. *Frequency* and *imaginability* were also predictive factors of the participants' choices. *L2 proficiency* was also found to play a role, with more proficient participants being more likely to produce translation options that were also preferred by native speakers.[23]

2.3.2 Effects of subsequent lexical language acquisition

There is a longstanding discrepancy between how much attention forward CLI has received, compared to reverse CLI. This has been, presumably, caused by an assumption that the L1 system of late bilinguals is relatively stable. In the present thesis, both forward and reverse CLI caused by translation ambiguity in the MML is investigated.

Some research that is relevant for the study of reverse CLI in multilingual speakers with respect to acquisition and processing of words with translation ambiguity has, however, been produced in the recent years. Some of this research has focused on CLI and some on attrition (see 2.2.2). The research has primarily focused on the effects of *cross-language activation* (CLA), *modulating factors*, and *permanence*, as well as explaining these. CLA refers to the effects of non-selective access (see Section 2.1) and has been found to cause both *facilitation* and *competition* effects. The former refers to positive (faster responses) and the latter to negative (slower responses) aspects of CLA. Working memory (WM), cognitive control (CC), and metalinguistic skills have been proposed as modulating factors. Finally, permanence refers to the longevity of the proposed effects.

In the case of CLA, both effects of activation and lack of activation of lexical items are expected to have an impact on lexical processing in the MML. Costa, Pannunzi, Deco, and Pickering (2017) have proposed that CLA of L1 and L2 words is a result of the way these words have been learned, and thus takes place primarily at early stages of L2 acquisition. From this perspective, we might see more effects of CLA at early stages of the acquisition (of particular individual words). It is possible, if not likely, that any early effects will primarily be present in processing rather than outcomes. The assumption of a discrepancy between

[23] It should, however, be noted that the other options produced – being actual potential translations of the given word – were not necessarily wrong *per se*, they just did not represent native target-like performance.

processing and outcomes is supported by Bice & Kroll (2015) who found that (very) early effects of reverse CLI from L2 Spanish to L1 English in a lexical decision task had no outcomes in accuracy or response times (RT) but could be observed in ERPs.[24]

Previous research of CLA has found both competition and facilitation effects. In general, bilinguals tend to be slower in word retrieval (than monolinguals), even in their dominant language (Gollan, Montoya, Cera & Sandoval, 2008). Meuter and Allport (1999) as well as Misra, Guo, Bobb, and Kroll (2012) found that there is a larger switching cost from the less dominant language to the more dominant language (than vice versa). The assumption, then, is that the longer RTs are related to inhibition of the dominant language, and that the latencies relate to the amount of inhibition applied (Higby, Donnelly, Yoon & Obler, 2020). Green's (1998) Inhibitory Control Model would suggest that interference (representing competition effects) takes place in bilinguals due to CLA in words that refer to the same concept. Thus bilinguals would need to deal with the resulting CLA by inhibiting those competitors that are not used for the task. On the other hand, representing results that are opposite to the predictions of the Inhibitory Control Model, there have been findings that translation equivalents facilitate word retrieval (Costa, Miozzo & Caramazza, 1999) representing faster RTs.

Higby, Donnelly, Yoon and Obler (2020) attempted to address the aforementioned discrepancy in the findings where both *competition* and *facilitation* effects have been found. Controlling for word frequency, word length, and whether the participants knew the translation equivalent (consciously), a facilitative effect was found for translation equivalents in the L1 – without presenting the L2 translation equivalents in any way. That is, when the participants knew a word in the L2, they were also faster at naming that word out loud in the L1. The authors hypothesize that this facilitative effect is caused by repeated parallel co-activation of translation equivalents raising their resting levels for both items (i.e., the target and its translation equivalent). The authors suggest that the facilitated lexical retrieval caused by use of words in a particular language represents a *direct frequency effect*, while the effect of use of the translation equivalents in another language has an *indirect frequency effect*.

WM, CC, and use of metalinguistic knowledge have been proposed as modifying factors of reverse CLI caused by CLA. Link, Hoshino, and Kroll (2008)

[24] ERP refers to *event related potentials*, which is a non-invasive technique for measuring brain activity (Luck, 2014).

tested the modulating effects of CC and WM. It was found that both greater WM and better CC relate to a reduction in CLA. Furthermore, the results suggested that WM would be more important in comprehension, whereas CC would play a larger role in production. As noted earlier, Jiang (2002) has proposed that tasks that allow the use of metalinguistic skills would probably show different results of CLI. Naturally, what falls under CLI and what falls under attrition is difficult separate in production data. However, effects in line with Jiang's proposals were found by Schmid & Jarvis (2014). What was hypothesized to be effects of L1 attrition affected free speech more than performance in tasks where the participants had a possibility to use metalinguistic skills.

Schmid & Jarvis (2014) suggest that there are two possible explanations for (what the authors lump under) attrition phenomena in the L1 that are related to relocation to a non-L1 environment: 1) that lexical access issues relate to increased need for regulation as a result of managing two (or more) linguistic systems, and 2) an increase in activation thresholds as a result of lack of use. As suggested by the Activation Threshold Hypothesis (ATH) items that are not used will be attrited (Paradis, M., 2007). A third explanation is offered by Levy, McVeigh, Marful, and Anderson (2007), who tested retrieval practice paradigm (Anderson, Bjork & Bjork, 1994, see retrieval-induced forgetting in Section 2.1) and found that using L2 words impaired their L1 equivalents. The explanatory power, at least for initial effects, of RIF was questioned by Runnqvist and Costa (2012) who failed to find an effect of RIF in the L1 as a result of L2 picture naming.

In line with the first explanation, Higby et al. (2020) hypothesized that immersion in the L2 environment causes regulation at a global level (rather than item level) due to the primarily L2 context causing slower access to L1 items globally. This, in combination with the boost for the items whose translation equivalents are actively used in the L2, would result in a generally slower L1 lexical access but with an advantage for the translation equivalents. What supports Higby and colleagues' hypothesis is that support for *temporary* (as opposed to *permanent*) nature of effects in the L1 have been found using qualitative means (Opitz, 2013), quantitative methods (Linck, Kroll & Sunderman, 2009),[25] and computational models (Meara, 2004, 2006).

[25] Baus, Costa, and Carreiras (2013) found effects of reduced availability of L1 lexical representations as a result of immersion (6 months) in picture naming but not in a *semantic fluency task* (i.e., list as many members of a category). Changes in L1 lexical access in the picture naming task were modulated by frequency and cognate status.

What falls under CLI and what under attrition is not completely clear based on the definitions of attrition. Ecke (2004) includes "the decline of any language (L1 or L2), skill or portion thereof in a healthy individual speaker" as attrition (p. 322). Schmid and Köpke (2008) highlight the existence of a L2 in, that "first language (L1) attrition refers to a change in the native language system of the bilingual who is acquiring and using a second language (L2)" (p. 210). Pavlenko (2004) highlights the importance of not interpreting the existence of CLI as evidence of attrition. She proposes five processes of interaction between languages: *borrowing*, *restructuring*, *convergence*, *shift*, and attrition. Note that in Pavlenko's classification, even attrition is subsumed under CLI. The proposal is that the first four types *can* represent attrition, but that should not be the assumption.

I will take a somewhat restricted stance in the present dissertation with respect to causality in attrition. The presumption is that individual representations in any language of a speaker can attrite as a part of regular processes of memory and consolidation. As pointed out in Section 2.1, when analyzing the results of the four empirical studies in the present dissertation, attrition will only be used to refer to effects caused by *lack* of activation, i.e., those falling under decay and aligning with the ATH. Any effects of CLA, irrespective of whether they are instantaneous or gradual, are lumped under the umbrella of CLI.

2.4 Summary and empirical studies

Both quality and quantity of CLI in the MML have been proposed to be affected by a range of modulating factors. Proficiency in each of the learner's languages is one factor. This dissertation therefore takes a developmental focus. Working memory and cognitive control have been proposed to influence both acquisition outcomes and lexical processing. In this thesis, these are subsumed under aptitude. An area that has seen much research in CLI is typological distance. This has been treated as genealogical relationship, which is always constant, and psychotypology, which can have between-subject and within-subject variation. While there is arguably much overlap between the two, psychotypology is measured in this thesis. Research on CLI has often focused on overt effects, either resulting in positive or negative transfer. In this thesis, though, the focus is on gradual changes rather than snapshot outcomes. Lexical processing is subject to both 1) instantaneous processes of priming caused by spreading activation and non-selective access, and 2) gradual changes caused by attrition and RIF. In this thesis, the manipulations are done in the experiments with respect to the former to tap into the latter.

The Parasitic Model (Hall & Ecke, 2003) proposes that factors that modulate CLI in the MML can be divided into five areas: *learner factors, learning factors, language factors, event factors, and word factors* presented in Table 5 below. The empirical work in this thesis covers aspects of all these five areas. Some of these aspects are controlled in the sense that they are kept constant. Other aspects are measured, meaning that there is a specific component in the experiment designed to capture data on this aspect. Finally, some aspects are manipulated, meaning that experimental stimuli has been designed to vary in these aspects.

Table 5

Controlled, measured, and manipulated aspects affecting CLI

AREA	FACTOR	Study 1	Study 2	Study 3	Study 4
Learner	Psychotypology		Measured	Measured	Measured
	Aptitude	Measured	Measured	Measured	
Learning	L2 status		Controlled	Controlled	Controlled
	Proficiency	Measured	Measured	Measured	Measured
	Acquisition order	Controlled	Controlled	Measured	Measured
Language	Typology	Manipulated	Manipulated	Controlled	Controlled
Event	Language mode			Controlled	Controlled
	Task	Measured	Measured	Measured	Measured
Word	Form	Controlled	Controlled	Controlled	Controlled
	Concept	Manipulated	Manipulated	Manipulated	Manipulated
	Concreteness	Controlled	Controlled	Controlled	Controlled
	Frequency	Controlled	Controlled	Controlled	Controlled
	Competitor frequency			Controlled	Controlled

All four experiments in this thesis have been designed to tap into the speaker's (un)conscious representations and processing, and how these change over time. The dependent variable that is central to this dissertation is variation at the level of concept: conceptual cross-linguistic similarity is manipulated in all four empirical studies. The independent variables that are central to this dissertation are psychotypology (Study 2-4), aptitude (Study 1-3), and proficiency (Study 1-4). The first two studies represent experimental data from artificial language learning.

Participants in these studies are complete beginners in the L2 Kontu v1 or the L3 Kontu v2. The latter two studies represent data from naturalistic learners with German L1, English L2, and Swedish L3. In Study 3, longitudinal data from CEFR A1 to C1 is presented. Study 4 covers the same proficiency range, but with cross-sectional data and proficiency in the three languages as a predictor.

3 Cross-Linguistic Influence in Early L2 Word Learning

3.1 Introduction

This dissertation is about cross-linguistic influence (CLI) in the developing multilingual mental lexicon (MML) in adults. The study reported in this chapter is about lexical CLI at the very initial stage[26] in second language learning. The initial stages are interesting given the astonishing feats language learners are able to perform already in the very beginning of the learning experience. For example, Gullberg, Roberts, Dimroth, Veroude, and Indefrey (2010) found that adult Dutch initial-state learners of Mandarin Chinese were able to implicitly extract information on segmentation, form-meaning mappings, and phonotactics from a seven-minute[27] weather report in Mandarin Chinese and apply that to novel items.

To date, the principal assumption has been that, in the early stages, *formal effects* – such as switching to another language, or attempting to construct words based on lexical items in previously acquired languages – of CLI dominate. On the other hand, at later stages of learning, *meaning-based effects* – such as semantic extensions like transfer of polysemy or homonymy to the target language (TL)[28] – of

[26] In the literature on early stages of post-L1 language acquisition, the term *initial state* is often used. Given the static nature of the term 'state,' and the developmental focus of this dissertation, the use of 'initial state' is reserved to the very moment at which language acquisition begins. For any broader references to periods, no matter how early in the learning experience, that include some form of learning or development, the term *stage*, for example *early stages*, will be used.

[27] Half of the participants watched the report two times, resulting in 14 minutes of exposure.

[28] As pointed out in Section 1.1, a *target language* (TL) in this dissertation is *any* language in which participant performance is presently tested irrespective of order of acquisition, i.e., the language *in which* CLI is observed. Given the multidirectional focus of this thesis, that means that the TL can be the L1, the L2, L3, or the Lx. *Source language* (SL), then, is the language *from which* CLI originates in any particular direction of interest. This means that the SL can also be any of the learner's languages irrespective of order of acquisition.

CLI dominate (Bardel, 2015; Ringbom, 2007). These findings might, to some extent, be a product of production error-focused sampling methods. Since the focus in this dissertation is primarily on the development of meaning-based effects – and to account for both conscious and unconscious lexical activity – the present study has been designed to tap onto both learning outcomes (i.e., overt effects) and processing (i.e., covert effects) in forward CLI in the L2. Furthermore, reverse CLI is investigated in the same population.

Forward CLI refers to the influence of a previously acquired language observed in a later acquired language. That means that when investigating forward CLI, the *source language* (SL) in the present study is the L1 English and the TL is the L2 Kontu v1 which is a pseudolanguage. *Reverse CLI* refers to the influence of a later acquired language observed in the chronologically previously acquired language. That means that when investigating reverse CLI, the SL in the present study is the L2 Kontu v1 and the TL is the L1 English.

The Parasitic Model (PM; Hall & Ecke, 2003) would predict immediate effects (both overt and covert) of forward CLI since new words in the L2 are presumed to be connected to L1 *host representations* (see Section 2.2.2). The host representation is the L1 word that is determined to be the closest match – either based on form, if possible, or meaning, in which case the L2 form connects to the conceptual representation via the L1 frame. Later, differences between the two languages are detected at the item level, presumably mitigating CLI. Furthermore, the learners will have to acquire skills to regulate *cross-language activation* (CLA; see Section 2.1 for accounts of *priming*, *spreading activation*, and *non-selective access*).

In addition to learning to deal with regulating CLA in forward CLI, suggestions have been made that the process of *L1 attrition*[29] starts already at the early stages of SLA (Schmid & Köpke, 2017). If speakers of multiple languages use the same, shared, conceptual system for their languages (c.f., Hall & Ecke, 2003; Jiang, 2000; Kroll & Stewart, 1994; Pavlenko, 2009), CLA of the lexical competitors is expected to cause slower word retrieval in multilingual speakers. On the other hand, it has been found that L2 word knowledge can have a facilitative effect on L1 word processing (for cognates and non-cognates alike), even in tasks conducted purely in the native-language (Higby, Donnelly, Moon & Obler, 2020). In the

[29] Note that Schmid & Köpke group both 1) effects caused by *lack* of activation, as well as 2) effects caused by the *addition* of the L2 under the umbrella of L1 attrition. See Section 2.3.2 for discussion.

present thesis, effects that are caused by CLA, irrespective of whether they are instantaneous or gradual, are lumped under the umbrella of CLI.

The observation that a speaker does not make overt errors does not necessarily imply that the underlying representations in the MML are not affected. It can be hypothesized that the effects of CLI – both forward and reverse – appear from the very moment the learner starts acquiring a new language, even if the reverse effects (i.e., from the L2 to the L1) are initially so subtle that they are difficult to measure. To maximize the potential of capturing early effects of reverse CLI, if present, the present study has been designed to map onto unconscious processing in the L1 (while for forward CLI, both conscious and unconscious processing are of interest).

In terms of methodology, to simulate the effects of limited vocabulary and early language learning, a pseudolanguage 'Kontu v1' was created and taught to the participants in the present study. Kontu v1 consists of 32 words with 36 meanings and is phonotactically adherent with Finnish. Twenty eight words had a clear dominating L1 translation equivalent and four forms had ambiguous meanings corresponding to two translation equivalents in the L1. This results in 36 meanings.

The acquisition of Kontu v1 words was tracked throughout the learning, as was forward and reverse CLI. Furthermore, the effect of aptitude was an object of interest in the study.

3.2 Aim and predictions

The aim of the present study was to investigate forward (from L1 English to L2 Kontu v1) and reverse (from L2 Kontu v1 to L1 English) CLI and the effects of aptitude at the very initial stages of second language acquisition. The participants in the present study are *ab initio* monolinguals and can be considered initial state learners at the onset of their participation in the study. Given these aims, the following research questions have been formulated:

RQ1 If present, is CLI at the initial stage in the MML unidirectional (only forward) or multidirectional (both forward and reverse)?

RQ2 To what extent is CLI in the MML affected by aptitude?

The first research question relates to the effect of stage of acquisition in directionality of CLI. Both potential directions are of interest, i.e., from L1 English to the L2 Kontu v1 and vice versa. Proficiency at the initial stage is operationalized as increased knowledge in the L2 Kontu v1. The assumption is that at this stage of learning, CLI will likely be disproportional in that it takes place mostly from the L1 to the L2. The second research question relates to the effect of aptitude, which is one of the modulating factors of interest in this thesis. Aptitude in the present study is operationalized as between-subject variation in collected measures of working memory (WM) and cognitive control (CC).

With respect to the early stage processes in CLI, lexical processing, and acquisition, explicit predictions (detailed in Table 6 below) are made by the Parasitic Model (PM, presented in Section 2.2.2). The PM is about L3 acquisition and has been adjusted to L2 acquisition for the purposes of the present study. As shown in Table 6 below, four processes are hypothesized to take place at the first stage of the PM.

Table 6

First stage of the Parasitic Model of Lexical Development (Hall & Ecke, 2003)

A1	A2	A3	A4
The L2 word form is registered in STM and the closest matches in L1 are activated based on salient form attributes.	The L2 form is connected to a host representation and is established in LTM in distributed fashion.[30]	Differences between L2 form and host representation are detected, new patterns are rehearsed and the representation is revised with respect to the attributes that distinguish it from the host and/or other consolidated neighbors.[31]	If no matching representation is activated sufficiently, the L2 form is connected to the frame of the nearest conceptual equivalent.

Note. STM is short-term memory and LTM long-term memory (adapted from Ecke, 2015).

[30] The *host representation* is normally the most highly activated related L1 form, where some threshold level of similarity between the items is met. *Distributed fashion* refers to the activation of same nodes in the network as the host form.

[31] Not always achieved, leading to fossilization of the interlanguage configuration (cf. Jiang, 2000).

The proposed timeline of the PM would suggest that Stage A1, where the L2 word form is attached to its closest match in the L1, takes place during the learning phase in the present study. The processing of meaning in the L2, Kontu v1, is expected to take place via the L1 English which means that we would expect effects of forward CLI already during learning. Effects of translation ambiguity are not expected in the (unlikely) case that the learners rely solely on short-term memory (STM) without any activation of linguistic knowledge. If the learners – in addition to the STM – use L1 resources, we would expect effects of translation ambiguity in the form of longer times on task (ToTs) in learning and assessment, and lower accuracy in assessment. Furthermore, the Inhibitory Control model (Green, 1986, 1998) would predict interference effects of regulating the L1 (see Section 2.1.2 for overview and Section 2.3.2 for specific suggestions for reverse CLI). An interesting empirical question is whether the learning of L2 Kontu v1 is able to cause effects from Stage A3 already during the learning despite Stage A2 having not yet taken place. If the learners actively detect the translation ambiguity between the L1 and the L2, they can either rehearse activation patterns – as suggested by the PM for Stage A3 – or attempt to deal with the ambiguity using introspection (see 2.3.1) and conscious rules (Jiang, 2002). In the case of the former, the effect of translation in ToTs, as compared to the development of items without translation ambiguity, is expected to reduce with increasing proficiency, whereas in the latter case the effect would be expected to sustain at least up until consolidation.

Based on the PM, Stage A2 would be expected take place during consolidation. If representation in a distributed fashion is not expected until consolidation, we should not observe any priming effects until after consolidation. Alas, any effects of translation facilitation (Costa, Miozzo, & Caramazza, 1999) would also not be expected to present themselves until at this stage.

Given the presented predictions from the PM as well as empirical insights (e.g., Elston-Güttler, 2000), the assumption is that for forward CLI the participants will be dependent of their pre-existing lexical-conceptual representations. Based on the suggestions of the PM and the L1 Lemma Mediation hypothesis (Jiang, 2000, 2002), this should particularly be the case at the early stages. On the other hand, we can hypothesize that at the very initial state, learners – and particularly those with limited experience from language learning – *will* assume between-language conceptual equivalence until there is evidence for the opposite. This means that, with respect to forward CLI, the assumption is that CLI *will* take place and that processing *is* affected by the L1-derived lexical-conceptual representations. The question for forward CLI, then, is primarily not *whether* there is forward CLI, but rather *how* it operates from the perspectives of accuracy and processing at the very initial stages.

For reverse CLI, the situation is different. There are suggestions (e.g., Schmid & Köpke, 2017) that reverse effects take place from the very beginning of second language acquisition. However, there is a shortage of both models making explicit predictions as well as empirical research about reverse CLI caused by between-language conceptual non-equivalence at the very initial stages. Given the lack of specific predictions, the question in the present study is rather *whether* any reverse CLI can be observed at the initial stages.

Finally, the effects of working memory (WM) and cognitive control (CC) are of interest in the present study as both have been postulated to have an impact on the acquisition and processing of vocabulary (Ellis & Sinclair, 1996). Available WM resources would be expected to facilitate learning whereas available cognitive resources in the form of CC should aid the learners in regulating the L1 (and presumably the L2, should it require any regulation), potentially leading to less CLA and, subsequently, less CLI.

3.3 Method

3.3.1 Participants

The participants were English L1 speakers with limited knowledge (≤ CEFR A2) of other languages. Thirty participants took part in the study. The participants' mean age was 23.89 (Min-Max: 18-41, SD = 6.78), and 73% were female. All had grown up in the Inner Circle countries of the English diaspora.[32] Data collection took place in Birmingham, Nottingham, and Lund. Most participants were studying a variety of different subjects at either the University of Birmingham, the University of Nottingham, or Lund University at the time of testing. Sixty-three percent of the participants were undergraduate students, whereas the rest were

[32] The *Inner Circle* refers to the first diaspora in Braj Kachru's (1997) model of World Englishes and includes United Kingdom, United States, Australia, New Zealand, Ireland, Canada, and South Africa.

graduate students or early-career staff. The participants in Lund all studied at English-medium programs and conducted their daily lives in English.

The participants filled in a basic two-page language background questionnaire (see Appendix A), which was developed based on the LHQ 2.0 questionnaire by Li, Zhang, Tsai, and Puls (2014) as well as Suhonen (2015). The questionnaire was administered in a pen-and-paper format before any further data collection. Responses in the questionnaire were used as exclusion criteria. Data was collected on two days with a single-night consolidation period in between. An additional questionnaire (see Appendix B) was filled in at the beginning of the second data collection session with the purpose of gathering data on the participants' activities between the two sessions with a focus on evaluating the participants' quality of sleep. Some of the questions were based on Kato (2014).

Despite limited experience in language learning beyond their L1, the participants were asked to self-evaluate their language learning aptitude (M = 4.13, Min-Max = 2-7, SD = 1.17), attitude (M = 5.07, Min-Max = 2-6, SD = 1.11), and ability for cross-cultural communication (M = 4.23, Min-Max = 1-7, SD = 1.55) on a seven-point Likert scale. The participants slept between the two data collection points and self-reported on average 7.33 hours of sleep (Min-Max: 6-10, SD = 0.91). They also self-reported their sleep quality (M = 5.26, Min-Max: 3-7, SD = 1.00) on a seven-point Likert-scale.

3.3.2 Instruments

3.3.2.1 Teaching Kontu v1 and measuring forward CLI

In the experiment, the participants were taught a pseudolanguage called Kontu v1 which was designed for the purposes of the experiment. During acquisition, the participants alternated between learning and assessment tasks. Since the participants acquired novel items, forward CLI is operationalized both as accuracy in assessment as well as time on task (ToT)[33] in learning and assessment. Forward CLI was measured throughout acquisition. The learning tasks were designed to maximize learning outcomes securing the best possible prospects for CLI to take

[33] Measures in *strictly* timed tasks where the participants were forced to operate as fast as possible limiting possibility for *introspection* are referred to as 'response time' (RT) in this dissertation while tasks where the participants were allowed to take their time, allowing introspection, are referred to as 'time on task' (ToT).

place. Explicit instruction was used, and the participants were motivated by using gamification and reinforcement. Spaced repetition continued past successful learning outcomes. Between data collection session one (where learning took place) and data collection session two, there was a one-night consolidation period.

The word forms in Kontu v1 consisted of 32 word forms taken from Finnish. Sixteen forms were bisyllabic and followed a CVCV (consonant-vowel-consonant-vowel) structure, and the remaining 16 of the forms were also bisyllabic but followed a CVCCV structure. No forms had letters that do not exist in English. Twenty eight of the word forms had a clear dominating translation equivalent in the L1. In addition, four word forms had translation ambiguity in the sense that they had two translation equivalents in the L1. This gives us 36 meanings: 28 forms with one unitary meaning (28 x 1), and four with two meanings (4 x 2). All translation equivalents were primary referents of nouns that could be easily depicted using pictures. Thus, 28 of the Kontu v1 forms were transparent in the sense that they could be translated into the participants' L1 without any translation ambiguity. Examples of this type of items are presented in Table 7 below.

Table 7

Form-meaning mappings in Kontu v1 without translation ambiguity

	FORM	MEANING
ITEM	lukko	logo
	kelo	receipt
	tuppi	frog
	huti	door

Note. None of the Kontu v1 forms were words in English and none had letters that do not exist in English.

Of the four forms with translation ambiguity, half were CVCV-forms and half CVCCV-forms. Participants were not informed about the translation ambiguity across the languages. Table 8 below presents the four Kontu v1 words that have translation ambiguity. A full list of the form-meaning mappings is available in Appendix C.

Table 8

Form-meaning mappings in Kontu v1 with translation ambiguity

	Form	Meaning 1	Meaning 2
	koti	flag	ticket
Item	talo	snake	queue
	lappu	bike	wheel
	pullo	paint	goal

Note. The Kontu v1 forms were taught with both meanings and with no indication to the participants that some of the items would have more than one translation equivalent in the L1.

The form-meaning mappings from Kontu v1 were taught using a combination of visual and auditory stimulus. Each form-meaning *learning sequence* consisted of a picture of the referent followed by a visual presentation of the word form below the picture. The learning phase was divided into four *blocks* with a break between each session. Learning sequences and blocks are illustrated in Figure 7 below.

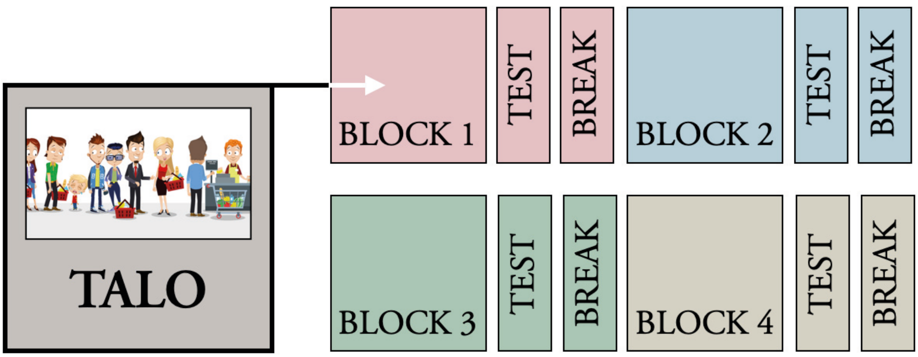


Figure 7. Learning sequences and blocks in Study 1

Each learning sequence consisted of a picture that illustrated a meaning and a Kontu v1 word form. Each block consisted of 4 x 36 learning sequences each representing a novel picture-form combination. There were a total of four blocks with identical (re-randomized) content. Each block was followed by a learning outcome test and a break. First block included auditory item presentation.

The blocks were identical with the exception that the order of the learning sequences was randomized separately for each block and participant. The four blocks consisted of a total of 576 learning sequences (4 x 36 = 144 learning sequences in each block). Each of the four instances of form-meaning pairs in a block included a different picture. Additionally, to improve learning as well as to make it more naturalistic, the first block included an auditory presentation of the word two times in the learning sequence. The participants were not required to complete the auditory stimulus but were free to move forward at their own pace. ToTs for each learning sequence were recorded, which allows for comparison across the items with and without translation ambiguity across the two languages. The inclusion of the auditory stimulus likely increases the ToTs in the first four presentations preceding the first learning outcome test.

A *learning outcome test* (LOT) was administered at the end of each block. This allows us to analyze the learning results following each block both in terms of accuracy and ToTs. In the LOT, the participants were presented with 36 picture selection tasks – one for each form-meaning pairing in Kontu – each with four pictures marked with letters from A through D. The participants were supposed to point out the correct form-meaning mapping by pressing buttons marked with A through D on the response pad. All form-meaning mappings were tested.

In the first three LOTs, feedback was provided in the form of a reinforcing buzzer sound for incorrect response prompting the participant to pay attention to the particular form-meaning pairing in the next block. No feedback for correct responses was provided. Furthermore, no feedback was provided in the fourth LOT which would have prompted the participants to pay attention to items they had not fully acquired, thus affecting priming data in the immediately following conceptual priming task. An example item is provided in Figure 8 below.

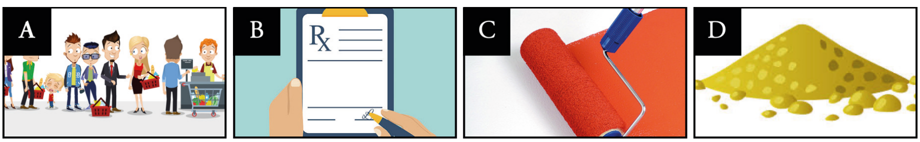


Figure 8. The learning outcome test

In the learning outcome test (LOT) the participants had to choose the correct meaning for a given word out of four options using a button box with buttons labeled A-B-C-D. The figure presents an example item. Each learning assessment component included 36 such items.

Learning outcomes were also tested after the participants had a one-night consolidation period. This served two purposes: 1) to ensure the participants had acquired the experimental items beyond STM, and 2) to see whether the processing of the items was affected by the consolidation period. The former was measured by looking at the participants' response accuracy and the latter by analyzing the participants' ToTs.

3.3.2.2 Measuring reverse CLI

In the present experiment, reverse CLI is operationalized as processing cost and priming in response time (RT) data when comparing performance immediately before the learning of Kontu v1, immediately after learning Kontu v1, and after a one-night consolidation period. A conceptual priming task was used for this purpose. The choice of this task was based on the assumption that increased connectedness in the associative lexical networks, into which one can tap with semantic priming tasks, would lead to either 1) increased priming between the two associated meanings taught in the learning phase, or 2) interference caused by competition effects slowing down word retrieval. A masked priming variant (Forster & Davis, 1984) was used.

The stimulus consisted of 52 target words. Thirty six of these were English translation equivalents of the 36 meanings from Kontu v1 (see Tables 7 and 8 above for examples). An additional set of 16 pseudowords was created – all of which phonotactically legal in English – using Wuggy (Keuleers & Brysbaert, 2010). The sole purpose of the pseudowords was to convince the participants that they were supposed to focus on whether the targets were real words or not. The pseudowords were matched in both syllable quantity and length in letters to the English forms of those items that were translation equivalents of the translation-ambiguous Kontu v1 words. The proportion of pseudowords as targets in the trials, distracting the participants from the true purpose of the task, was 20%.

The primes consisted of 1) the translation equivalents of those words that had translation ambiguity between the L1 English and the L2 Kontu v1 in the 'kontu' condition acting as primes for other words from the same condition, 2) either related words that were also translation equivalents of other Kontu v1 words or unrelated words matched for length and roughly for frequency in the 'baseline' condition – these were distributed equally, 3) non-words in the 'pseudo' condition, and 4) translation equivalents of Kontu v1 items or non-words in the 'distractor' condition. The proportion of related prime-target combinations was 42%. Appendix D provides a complete list of prime-target combinations.

The participants responded to the lexical stimulus in the L1 English using a button box. They were asked to judge – as fast as possible – whether the displayed item was a word in English. The same single-language prime–target combinations from a pre-test (preceding the learning of Kontu v1) were displayed in a post-test and a delayed post-test. As such, the pre-test functions as a baseline to evaluate the change of the priming effect and processing cost of the experimental items with translation ambiguity in relation to baseline and filler items. The presentation order of the prime–target combinations was randomized.

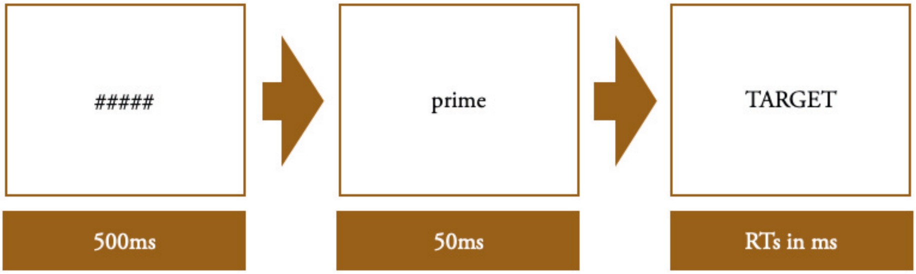


Figure 9. Sequence of presentation in the priming task

The stimulus (Figure 9 above) was presented using a so-called three-field paradigm (mask-prime-target) with the forward mask being presented for 500ms and the prime for 50ms. The interstimulus interval was zero milliseconds. The forward mask was a row of hash marks, the prime was presented using lower case letters, and the target using capital letters. The stimulus was displayed centered on the screen using SuperLab 5.[34] The participants responded to the stimuli using two color-coded (red = no, green = yes) buttons using a Cedrus RB-740 response pad.

3.3.2.3 Measuring aptitude

A type of continuous performance task, namely the *n-back task* (Kirchner, 1958), was used as a control measure for working memory[35] and the closely correlated fluid intelligence.[36] Both two-back and three-back variants with visual presentation of the stimuli were used. In the *n*-back task, the participant has to rapidly

[34] Superlab is a software by Cedrus Corporation. The response pad which was used for data collection was also from Cedrus. The Cedrus RB-740 response pad has seven buttons.

[35] *Working memory* is a cognitive system that holds information available for processing.

[36] *Fluid intelligence* refers to the capability of solving novel problems in Cattel's (1971) two-structure division of general intelligence.

decide whether the currently presented stimulus is the same as the one that was presented *n* turns back. This is outlined in Figure 10 below.

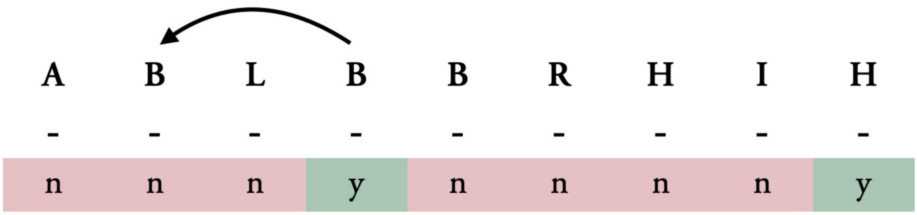


Figure 10. Stimulus in the *n*-back task of working memory

The stimuli cycle above presents an example of a 2-back task. Each letter is presented individually and the participant has to recall whether the presented letter is the same as the one presented two trials ago. The first row has the presented stimuli and the bolded bottom row represents the correct response (n = no, y = yes). For example, in the fourth stimuli 'B' the participant has to recall whether the letter two trials ago was also 'B,' which is the case. In this trial the correct response is 'yes.'

The stimulus set consisted of 15 letters, presented individually. Each letter was presented for a maximum of 2,000 milliseconds and the next stimulus was presented either following a response from the participant or after 2,000 milliseconds from the onset of the previous stimulus, whichever occurred first. Visual feedback was provided based on the participant's response using two rectangular boxes: one above the stimulus location and one below the stimulus location on the screen. Green indicated correct response and red incorrect response. No feedback was provided in the cases the participant did not provide a response in time.

The participants were first presented with instructions for the *n*-back task followed by a two-back training session with 50 trials. The training session was followed by a two-back task with 50 trials and a three-back task with 50 trials. Stimulus presentation was administered using SuperLab 5. The participants responded to the stimuli using two color-coded (red = no, green = yes) buttons using a Cedrus RB-740 response pad.

In addition to the WM task, a cognitive control task was administered to the participants using the *Eriksen Flanker task* (Eriksen and Eriksen, 1974), where the participants were asked to press a button should a stimulus belong to one of two categories (see category-button pairings in Figure 11 below). The participants were presented with seven letters on the computer screen. Their task was to solely focus on the middle one. The other six letters were distractors. Each of the two categories was assigned with a button on a button box and participants were asked to respond using the buttons as fast as possible. In some of the trials, the participants were presented with *congruent stimulus* where all the seven letters belonged to the same category and in some of the trials the participants were presented with *incongruent stimuli* where the six distractor letters belonged to a different category than the middle letter. The *flanker effect* was calculated by deducting the average result of the congruent category responses from the average result of the incongruent category responses. Figure 11 below presents an example of a congruent and an incongruent trial. In the congruent trial, the participant would have to press the left button. In the incongruent trial, the participant would have to press the right button. In both cases the distractors are connected to the left button.

Category-button pairings

Left button	Right button
Letters X and C	Letters V and B

Examples of congruent and incongruent trials

Congruent	Incongruent
XXXCXXX	XXXBXXX

Figure 11. Stimulus in the flanker task

There are two categories with two letters each. These are presented under Category-button pairings. Examples of congruent and incongruent categories were presented. The participants were asked to focus on the middle letter. In the congruent example both the middle letter 'C' and the distractor letters 'X' belong to category 'Left.' In the incongruent example the middle letter 'B' belongs to category 'Right' and the distractor letters 'X' belong to the category 'Left.'

The participants did a trial round to get comfortable with the task before actual data collection and were able to ask questions before proceeding with the task. The experiment tests the participants' ability to ignore the distractor stimuli that surrounds the focus stimuli (in the center) that the participant is supposed to focus on. The smaller the difference is between the congruent and the incongruent conditions, the better the participant is at inhibiting distracting stimuli. Stimulus presentation was administered using SuperLab 5, and the participants responded to the stimuli using two buttons on a Cedrus RB-740 response pad.

3.3.3 Procedure

The participants were first informed about the purpose of the study, asked to sign a consent form, and given the opportunity to ask questions. After it was mutually agreed that data collection could commence, the participants were asked to fill in the background questionnaire. If the participants fulfilled the participation criteria, they took the priming pre-test. The participants had been pre-screened and no participants needed to be excluded due to their responses in the background questionnaire.

The whole experiment consisted of five main sections (visualized in Figure 12 below): the background section, the pre-test, the learning of Kontu v1, the post-test, and the delayed post-test. The aptitude tests (*n*-back and flanker) were administered in the final section, together with the delayed post-test. Throughout the procedure, two types of breaks were given. Between major sections, the participants were instructed to take breaks involving physical movement. The pre-test, the learning component, the post-test, the delayed post-test, and the aptitude tasks also included assigned micro-breaks where the participants were instructed to rest their eyes for a moment to ensure that they could concentrate efficiently.

BACKGROUND ENGLISH	Welcome Consent for participation[37] Background questionnaire
PRE-TEST ENGLISH	Priming task in L1 English (functions as a base-line)
LEARNING KONTU	Learning Kontu v1 (four learning blocks and four assessment blocks)
POST-TEST ENGLISH	Priming task in L1 English (identical to pre-test)
Break (one night)	
DELAYED POST-TEST ENGLISH	Sleep quality questionnaire Priming task in L1 English (identical to pre-test) Learning outcomes (one assessment block) Flanker (cognitive control) Working memory (*n*-back) Debriefing

Figure 12. Procedure in Study 1

[37] Both University of Nottingham and University of Birmingham performed an ethics evaluation and approval of the study. A privacy policy and data management description was provided to the participants. In accordance with GDPR, the European Union General Data Protection Regulation (2016/ 679), the participants consented in writing to the following: 1) that their personal information will be stored separately from research data, 2) that personal information will be stored in relation to their compensation in accordance with the Swedish accounting legislation (BFL 1999: 1078) by Lund University centrally, 3) the collected but anonymized research data will be publicly available and may be used for further research purposes without additional consent from the participant, and 4) that it is possible to retract one's participation up until the point of publication. The participants were compensated with movie tickets in Sweden and gift cards in the UK.

The pre-test functions as a baseline to evaluate the change of response times of the experimental items that are related in Kontu v1 in relation to the baseline and distractor items. The presentation order of the prime–target combinations was randomized for each participant. All priming tasks were introduced to the participants as decision-making speed tasks.

After the pre-test and a break, the participants took the learning module (wearing headphones), which was the longest section of the experiment consisting of four learning blocks and four LOTs. This was once again followed by a break, after which the participants were administered the same masked priming task as in the pre-test. The item order randomization was performed again for the post-test. The purpose of the test was to see whether the participants would be affected by the acquired Kontu v1 items in their L1 immediately after the (short but) intensive learning period. Generic learning effects of the items in the priming task can be expected since the test was administered within a short time period from the original test. These effects can be adjusted for by comparing the development of the different item types (with translation ambiguity, without translation ambiguity, distractor non-words, and fillers). All-in-all, the first data collection session took 60 to 75 minutes for most participants. At this point, the participants left the premises.

The participants returned the next day for the delayed post-test. Upon arrival, the participants filled in a sleep quality questionnaire. The delayed post-test included the lexical priming task (identical to the pre-test and the post-test), a LOT, and the aptitude components (WM and CC). Item order randomization was performed again for the delayed post-test. The second data collection session took approximately 20 minutes for most participants.

Finally a debriefing was performed. The participants were also asked whether they were explicitly aware of the multiple meanings associated with the four experimental forms as well as if they were aware of something intervening the forward mask and the target. Most participants reported some awareness of the translation ambiguity as well as being aware of some presentation of stimulus between the hashes and the targets. No participant reported having been able to consciously read the primes. Participants were actively informed about the primes, the purpose of the priming task, and that Kontu v1 is not the same as Finnish at this point.

3.3.4 Data analysis, transformations, and structure

There are four outcome variables in Study 1: 1) time on task (ToT) in the learning sequences, 2) ToTs in assessment, 3) accuracy in assessment, and 4) response times (RT) in the priming task. The first three represent forward CLI and the last reverse CLI. A detailed description of the distribution of the data in Study 1, before and after any trimming, is provided in Appendix E. All statistical analysis was performed in R version 3.6.0 (R Core Team, 2018) using R Studio version 1.2.1335 as the graphical user interface. Data exploration was done using Jamovi version 1.0.8.0 (The Jamovi Project, 2019).

The ToT data for learning sequences and assessment (variables 1 and 2 above) contained in total 22,386 observations. Lower-spectrum trimming was set to 100ms (n = 290) representing 1.28% of the data. This left 22,095 time observations. The two conditions were almost equally affected by the missing data and the transformations with 97.42% data left in the condition without translation ambiguity and 97.48% data left in the condition with translation ambiguity. Furthermore, one obvious unplanned break of 152s was removed. No further upper spectrum trimming was done since the primary purpose of the learning segments was to maximize the participants' learning of the items and as such time limits were not enforced.

In addition to the ToT data from learning and assessment, the data available on learning outcomes include binary data on the participants' scores in the form of correct and incorrect responses in the LOTs. These 5,400 observations have been annotated with the available background data as well as the associated learning data from the 16 occurrences of Kontu v1 words in learning in the form of ToTs. Overall accuracy in the four LOTs was extremely high, which indicates that the participants actively engaged in learning. No trimming or transformations were done to this data.

The RT data for priming tasks contained a total of 10,440 observations. Response times of <200ms and >1000ms were removed (n = 130) which represented 1.24% of the data. This left 10,310 response time observations. Distractor items were most affected (1.7%). Out of the items in the conditions of interest, 0.7% of items without translation ambiguity and 0.2% of items with translation ambiguity were affected.

In addition to the outcome variables, the data set contains data on individual differences in the sample population. The majority of these come from the background questionnaire (see Appendix A). Aptitude data consists of three types of data: self-evaluations, WM scores, and CC scores. No trimming was deemed necessary for WM or CC data as both tasks timed out. The WM or CC data contain 1,920 observations in total. Descriptive data on the participants' performance in the aptitude measures, including correlations between subjective and objective aptitude measures, is available in Appendix E.

Separate linear mixed effect models (LMEs) were performed for each outcome variable. For all models, random intercepts were included for PARTICIPANT and ITEM. The main (fixed) factors TIME and ITEM TYPE were added as interactions. WORKING MEMORY and COGNITIVE CONTROL were added as co-variates. TIME (increased proficiency) was operationalized as either progress in 1) learning over the 16 presentations of each item, 2) five learning assessment modules, or 3) the three time points in the priming task depending on the outcome variable. For the LME for reverse CLI, RETENTION (difference in performance in LOTs 4 and 5) was added as an additional co-variate. The models are presented in Tables 9, 11, 13, and 14 in the Results-section below.

3.4 Results

3.4.1 Forward CLI in L2 Kontu

Forward CLI in the present study refers to the effect of the L1 English on the L2 Kontu v1. There are three measures that can capture such effect: 1) processing cost (ToTs in learning across item types), 2) processing cost (ToTs in assessment across item types), and 3) accuracy (correct responses in assessment across item types).

A ToT on a specific learning sequence represents an estimation of difficulty of acquiring that particular form-meaning mapping. The development of these was tracked across item types over the course of learning. Figure 13 below shows the ToTs for the two types of stimuli available in the learning task throughout their sixteen occurrences.[38] The figure shows that the TOTs reduce successively from the first occurrences towards the last one. Some increase in ToTs can be seen after the LOTs likely as a result of noticing. The exception to that rule is ToTs after the first assessment module. Initially the time spent on items with translation ambiguity was higher. This effect fades rather quickly and is virtually non-existent

[38] Table E7 in Appendix E provides additional data on ToTs in the learning sequences.

towards the end of the learning period. The initial difference between the two item types could partially be explained by conceptual restructuring and forward CLI in Kontu v1 from English. When comparing the first and second instantiations of the same form (with different meaning), the second instantiation is longer on average. The estimated effect of 195ms is, however, not significant (p = 0.71), meaning that the analysis does not support the hypothesis.[39]

Figure 13. Time spent on task (ms) in learning sequences

The numbers 1-16 refer to the amount of times the participant has been presented with the given form-meaning pairing in Kontu v1. MONO refers to items with a single translation equivalent in the L1 and POLY to items with two translation equivalents in the L1. The numbers above the bars are means (in ms) across participants and items within the given category.

Each form-meaning pair was presented 16 times, but those forms that map to two meanings have been presented a total of 32 times: 16 times with one meaning and 16 times with another (unrelated) meaning.[40] The descriptive data seems to indicate that there is an effect of increasing proficiency on ToTs, but that after a few occurrences the time seems to stabilize to approximately 1,500 milliseconds.

A linear mixed effect model was performed to compare the development of processing cost over time during learning across the two item types. The model is presented in Table 9 below. There is an estimated effect of 125ms for translation ambiguity (labeled ITEM TYPE in the model output) but this effect is not significant. The only significant effect is the generic learning effect over time.

[39] Additional data on this analysis is provided on Table E4 in Appendix E.

[40] The reason behind this choice was that testing reverse CLI was one of the objectives and the occurrences for each form-meaning pair were matched at the meaning (rather than form) level.

Table 9

Forward cross-linguistic influence in processing cost over time in learning

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.1231	-0.3686	-0.0525	0.2537	27.4025

Random effects

GROUPS	NAME	VARIANCE	SD
ITEM	INTERCEPT	1815641	1347.5
PARTICIPANT	INTERCEPT	571035	755.7

NUMBER OF OBS: 16833, PARTICIPANT, 30; ITEM, 288

Fixed effects

	ESTIMATE	SE	df	t	*p*	
INTERCEPT	4112.126	169.847	59.320	24.211	< 0.001	***
TIME (INSTANCE)	-112.838	5.465	15229.34	-20.647	< 0.001	***
ITEM TYPE	125.316	210.049	336.595	0.597	0.551	
WORKING MEMORY	-39.223	148.809	26.997	-0.264	0.794	
COGNITIVE CONTROL	70.601	148.820	27.000	0.474	0.639	
TIME:ITEM TYPE	-9.333	11.583	15231.41	-0.806	0.420	

REML criterion at convergence 304078.4

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to first presentation instance, baseline items, mean working memory, and mean cognitive control. WORKING MEMORY and COGNITIVE CONTROL have been normalized using Z-transformation. Time (generic learning effect) is significant. Formula: rt ~ time * itemtype + scale(wm) + scale(cc) + (1|participant) + (1|event).

Response times were also recorded for the assessment modules. A difference between the two conditions would indicate CLI. The observations are presented in Table 10 below.

Table 10

Mean and median ToTs (ms) in assessment modules by item type

TEST	ITEM TYPE	MEAN	*SD*	MEDIAN
1	MONO	4036	2506	3312
	POLY	4592	3469	3549
2	MONO	2824	1713	2277
	POLY	3195	2202	2539
3	MONO	2416	1301	2054
	POLY	2509	1450	2084
4	MONO	2328	1298	1948
	POLY	2587	1446	2094
5	MONO	2372	1329	1978
	POLY	2320	1098	1846

Note. MONO refers to items with a single translation equivalent in the L1 and POLY to items with two translation equivalents in the L1. The participants took the test on five occasions. The test was taken immediately after each of the four learning modules. The fifth learning outcome test was administered after a consolidation period of one day. In tests one through four, the participants are slower with the POLY items. After consolidation, the participants are faster with the POLY items.

As Table 10 shows, the participants were overall slower in the items with translation ambiguity. These items seemed to require more processing effort from the participants all the way until after the consolidation period. The average response times in the learning outcome test decreases after each module, but the difference between the items with and without translation ambiguity persists until the fourth learning assessment module.

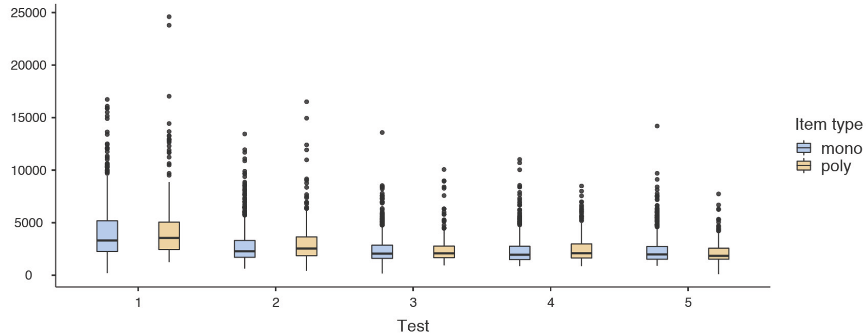


Figure 14. Time spent on task (ms) in LOTs. Item type refers to whether the Kontu form has one or two translation equivalents in the L1: items marked MONO have a single congruent translation equivalent in the L1 while items marked POLY have two translation equivalents. All times are given in milliseconds across participants and items.

What we can also note from Figure 14 above is that the items with translation ambiguity elicit a higher proportion of excessive time spent on task. One possibility is that these represent those instances when the participants consciously realize the existence of translation ambiguity, and thus these values were not removed in upper level trimming (see section 3.3.4). Once again, the mean ToTs for the first instances are higher for the items with translation ambiguity so an analysis was performed whether this difference can be explained by comparing the first instances of a form with first instances of a form with another meaning. There is an estimated effect of 674.6ms (p = 0.101).[41]

A LME was performed to compare the development of processing cost in assessment modules throughout learning across item types. The model is presented in Table 11 below. The main effects for TIME, ITEM TYPE, and WM are significant. Interaction between TIME and ITEM TYPE is also significant. The interaction indicates a presence of forward CLI.

[41] Table E9 in Appendix E presents more details about the analysis.

Table 11

Forward cross-linguistic influence in processing cost over time in assessment

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.8255	-0.5614	-0.1747	0.2916	13.2316

Random effects

GROUPS	NAME	VARIANCE	SE
ITEM	INTERCEPT	447991	669.3
PARTICIPANT	INTERCEPT	386864	662.0
RESIDUAL		2573547	1604.2

NUMBER OF OBS: 5252, PARTICIPANT, 30; ITEM, 72

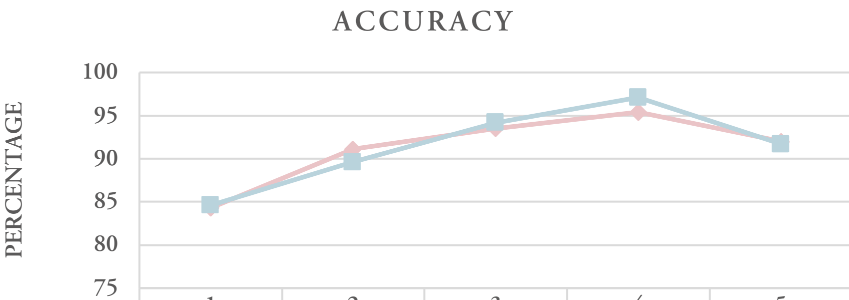
Fixed effects

	ESTIMATE	SE	df	t	*p*	
INTERCEPT	4589.32	179.62	122.30	25.550	< 0.001	***
TIME (BLOCK)	-573.37	32.03	1496.32	-17.997	< 0.001	***
ITEM TYPE	943.00	295.22	200.72	3.194	0.0016	**
WORKING MEMORY	-259.92	124.03	26.99	-2.096	0.0456	*
COGNITIVE CONTROL	-11.58	124.08	26.97	-0.093	0.9263	
TIME:ITEM TYPE	-227.84	68.00	1492.14	-3.351	< 0.001	***

REML criterion at convergence 92640.6

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to first assessment block, baseline items, mean working memory, and mean cognitive control. WORKING MEMORY and COGNITIVE CONTROL have been normalized using Z-transformation. Time, relationship, working memory, and interaction between time and relationship are significant. Formula: rt ~ time * itemtype + scale(wm) + scale(cc) + (1|participant) + (1|item).

The primary purpose of the learning component was to ensure that the participants would acquire the Kontu v1 words to a robust level. Based on the accuracy measures presented in Figure 15 below, this was indeed the case. The learning outcome assessment tasks included four options, none of which were novel to the participants. As such, chance would be at 25%. After the first four instances, the participants performed at 84.6% and 84.3% on the items with- and without translation ambiguity, respectively. This increased to 97.1% and 95.4%, respectively, by the fourth learning outcomes task, when the participants had encountered each form-meaning pairing sixteen times (see Figure 15 below). As we can see, there are no major differences in learning outcomes for the items with- and without translation ambiguity, which suggests that in terms of learning accuracy neither item type poses an issue for novice language learners.



	1	2	3	4	5
mono	84.3	91.1	93.5	95.4	92
poly	84.6	89.6	94.2	97.1	91.7

Figure 15. Accuracy in assessment modules. MONO refers to items without- and POLY to items with translation ambiguity. The participants did the test on five occasions. A test was taken immediately after each of the four learning blocks. The fifth learning outcome test was administered after a consolidation period of one night. In the post-test (4), accuracy for items with translation ambiguity is higher than for items without. This is likely associated with noticing. Accuracy in the delayed post-test (5) is overall lower than in the post-test and the drop for items with translation ambiguity is higher (at 3.4% for MONO and 5.4% for POLY).

The fifth learning outcome task took place after a one-night consolidation period. Attrition is represented in reduction of accuracy between the fourth- and the fifth learning outcome task. We can note that accuracy has been reduced to approximately 92% for each type of experimental items, which means that some attrition

(*M* = -0.038, Min-Max = -0.194 – 0.083, *SD* = 0.066) of the acquired information has taken place.

In addition to accuracy measures, the data contains self-reports about the participants' activities between the fourth and fifth assessment points including sleep quantity, quality, and consumed alcohol in standard units. Correlations between these factors and retention (attrition) are presented in Table 12 below.

Table 12

Correlation matrix between sleep quality, quantity, alcohol consumption, and retention

		RETENTION	SQUALITY	SHOURS	ALCOHOL
RETENTION	Pearson's r	—			
	p-value	—			
SQUALITY	Pearson's r	-0.086	—		
	p-value	0.650	—		
SHOURS	Pearson's r	-0.087	0.192	—	
	p-value	0.647	0.310	—	
ALCOHOL	Pearson's r	-0.252	-0.008	-0.131	—
	p-value	0.179	0.968	0.490	—

Note. RETENTION refers to the change in accuracy (0-1) between assessment point 4 and assessment point 5. Assessment 5 took place after one night's sleep. SQUALITY refers to sleep quality (self-reported), SHOURS refers to amount of sleep in hours (self-reported), and ALCOHOL refers to amount of standard units of alcohol consumed (self-reported).

All of the predictors have a negative correlation with retention, that is, increased sleep quality, quantity, and consumed alcohol all correlate with more attrition. None of these correlations is significant. The correlation between (self-reported) sleep quality and quantity with retention is minimal. The correlation for consumed units of alcohol and retention is higher. What is surprising, is that there does not seem to be any correlation between (self-reported) sleep quality and alcohol consumption.

A LME was performed to compare the development of accuracy in assessment modules throughout learning across the two item types. The model is presented in Table 13 below. While CLI was found for processing cost in assessment, CLI did not manifest itself in accuracy. This means that the participants spent more time arriving at the correct response for the items with translation ambiguity but nevertheless were equally accurate in their responses. The Kontu v1 learners are capable of reaching high accuracy in lexical items in the target language irrespective of the items' conceptual equivalence between the two languages.

For forward CLI, aptitude scores did not have any significant effects in the participants' performance on ToTs during learning. In the learning outcomes test, WM had a significant effect in processing (p = 0.46) while CC did not (see Table 11 above). Neither WM nor CC had significant effects in accuracy.

Table 13

Forward cross-linguistic influence in assessed accuracy over time

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-9.7530	0.1428	0.2100	0.3139	1.6886

Random effects

GROUPS	NAME	VARIANCE	SE
ITEM	INTERCEPT	0.6043	0.7774
PARTICIPANT	INTERCEPT	0.6340	0.7962

NUMBER OF OBS: 5393, PARTICIPANT, 30; ITEM, 72

Fixed effects

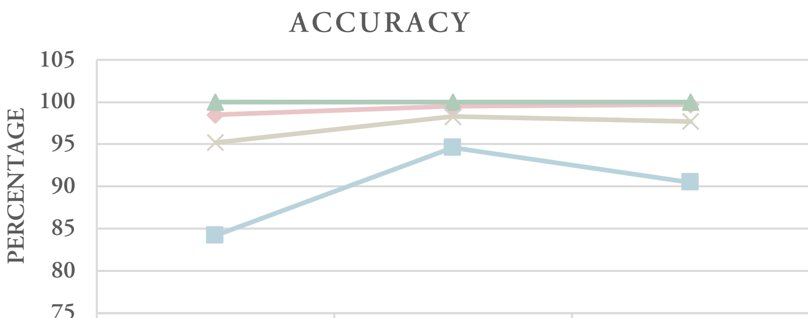
	Estimate	SE	df	t	p
INTERCEPT	1.91726	0.26594	7.206	< 0.001	***
TIME (BLOCK)	0.32939	0.06197	5.315	< 0.001	***
ITEM TYPE	-0.01399	0.46217	-0.203	0.839	
WORKING MEMORY	0.24122	0.16930	1.425	0.154	
COGNITIVE CONTROL	0.12082	0.17023	0.710	0.478	
TIME:ITEM TYPE	0.01136	0.12932	0.089	0.929	

Model fit AIC 2824.1 BIC 2876.9 logLik -1401.1 deviance 2808.1 df.resid 5385

Note. Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod]. Performed in R 3.6.0 using using lme4 (1.1-23) glmerControl (optimizer = "bobyqa"). INTERCEPT is set to first assessment block, baseline items, mean WM, and mean CC. WORKING MEMORY and COGNITIVE CONTROL have been normalized using Z-transformation. The only significant effect is time (operationalized by progress in assessment blocks). Formula: glmer(accuracy ~ time*itemtype + scale(wm) + scale(cc) + (1|participant) + (1|event), family = "binomial", control = glmerControl(optimizer = "bobyqa")).

3.4.2 Reverse CLI in L1 English

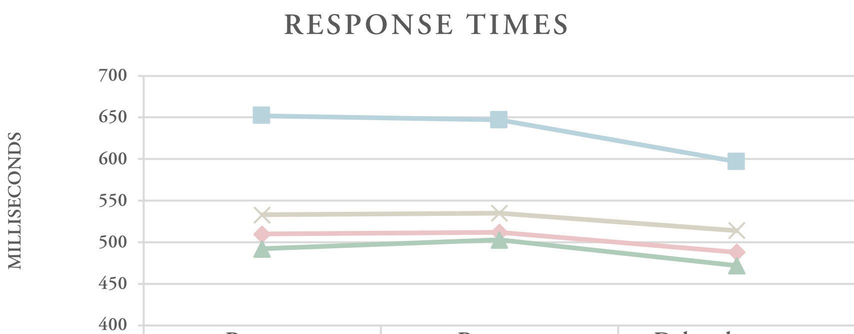
The second potential direction of CLI in Study 1 is from the L2 Kontu v1 to the participants' L1 English. Recall that four types of form-form pairings were presented in the masked priming experiment: baseline items that were English words whose translation equivalents were taught in Kontu v1, experimental items where the prime and the target are related in Kontu v1 but not in English, and two types of matching distractors ('dist' and 'pseu' below, neither is of interest in the analysis). The primary interest is the development of the experimental pairs that are related in Kontu v1 in relation to the baseline items. Some learning effects are expected for all items, including distractor items, since these are novel to the participants in their first instantiation. The analysis should, as such, be able to discriminate between generic learning effects and effects that are caused CLI.



	Pre	Post	Delayed
base	98.5	99.5	99.7
dist	84.2	94.6	90.5
kontu	100	100	100
pseu	95.2	98.3	97.7

Figure 16. Accuracy in the masked priming task. 'Base' to refers to baseline items that are translation equivalents of Kontu words with a single translation equivalent in the L1. 'Dist' refers to nonce words that are legal in English phonotactics. 'Kontu' refers to item pairs that are related in Kontu v1 but not in English. 'Pseu' refers to English words that are preceded by pseudowords in the mask. The participants took the test on three occasions: immediately before learning Kontu, immediately after, and after one-night consolidation period. The accuracy for all item types is high but the distractors items seem to cause most issues to the participants.

The participants' accuracy in deciding whether a particular item is an English word is presented in Figure 16 above. Distractor items seemed to cause most issues. Figure 17 below presents the RTs in the pre-test, post-test, and delayed post-test priming tasks. These were performed solely in the L1. No adjustments for generic learning effects have been made. The figure shows that the non-word target distractors are the most different from all the three other types of stimuli. Analysis of the effects of Kontu v1 on the learning data requires the data to be normalized using baseline items to account for general learning effects that are an effect of the test being repeated multiple times in the same experiment.



	Pre-test	Post-test	Delayed post-test
base	510	512	488
dist	652	647	597
kontu	492	503	472
pseu	533	535	514

Figure 17. Visualization of RT data. No adjustment for generic learning effects have been done. Data represents is mean response times in milliseconds. See Table E10 in Appendix E for more data.

Figure 17 above also shows that the learning effect appears primarily after the consolidation period for distractor items that were novel to the participants in their first instantiation. This indicates that early word learning can cause familiarity effects, but that those seem to require a consolidation period.

Both baseline items and distractor items in Figure 18 below have been centered to their averages at pre-test. Furthermore, the baseline is here kept constant through post-test and delayed post-test by deducting any learning effect at that level from both the baseline and distractor items.

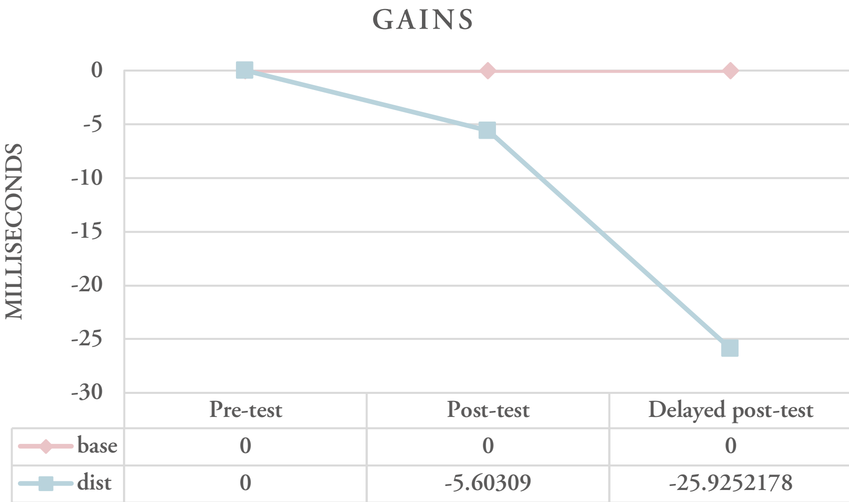


Figure 18. Visualization of data for assessing learning effects. Pre-test results are centered and baseline results kept constant throughout by deducting changes in those from all item types. This allows us to estimate the learning effect and consolidation effect for novel items.

In Figure 19 below, baseline items and items with translation ambiguity have been centered to their mean at pre-test. The baseline is kept constant through post-test and delayed post-test by deducting any learning effect at that level from both the baseline items and items with translation ambiguity. We can note that initially the RTs for those items increase. After the consolidation period the RTs decrease. This effect is not significant (see Table 14). The Kontu v1-derived items do operate in an unsurprising direction: in the immediate post-test there is presumably an increased processing cost for translation ambiguity representing either competition effects or RIF. After consolidation, an effect of priming or facilitation can be observed.

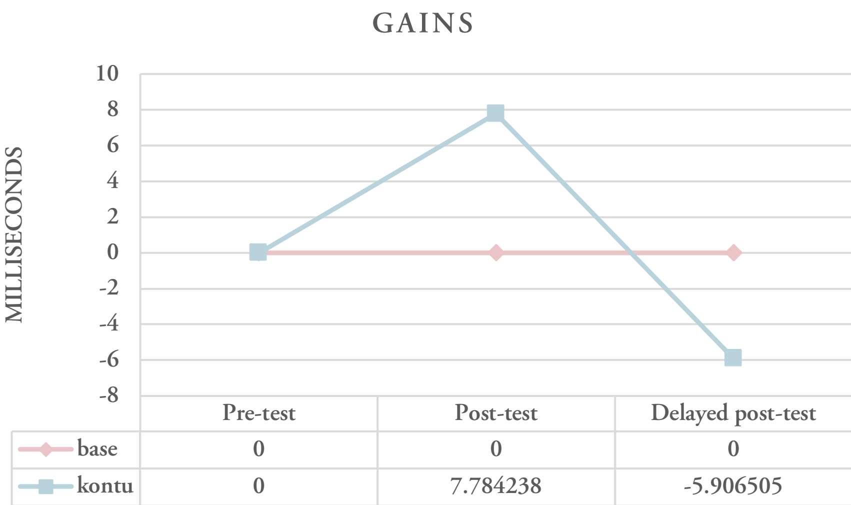


Figure 19. Visualization of data for assessing learning effects. Pre-test results are centered and baseline results kept constant throughout by deducting changes in those from all item types. We can see that Kontu items have higher response times (in comparison to baseline items) in the immediate post-test and decreased response times in the delayed post-test. This would align with the assumption of increased processing cost in the immediate post-test and conceptual priming after consolidation.

A LME was performed to assess reverse CLI from the L2 Kontu v1 to the L1 English. The model is presented in Table 14 below. The only statistically significant effect in the reverse CLI condition is a generic learning effect caused by the close proximity in time between the pre-test, post-test, and delayed post-test. There is an effect of increased priming (or facilitation) for items that are related in Kontu v1 as a result of learning but this effect is not significant. Consolidation, as operationalized by retention, shows no significant effects either. Also, there were no aptitude effects for reverse CLI.

Table 14

Reverse cross-linguistic influence in priming over time

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.6639	-0.6226	-0.1425	0.4301	6.0129

Random effects

GROUPS	NAME	VARIANCE	SE
ITEM	INTERCEPT	722.5	26.88
PARTICIPANT	INTERCEPT	2797.4	52.89
RESIDUAL		7116.1	84.36

NUMBER OF OBS: 5722, PARTICIPANT, 30; ITEM, 64

Fixed effects

	Estimate	SE	df	t	p	
INTERCEPT	525.6561	10.7758	38.8754	48.781	<0.001	***
TIME (BLOCK)	-11.1337	1.4602	5626.92	-7.625	<0.001	***
ITEM TYPE	-15.7001	13.5110	156.430	-1.162	0.247	
RETENTION	7.7583	9.9018	25.9962	0.784	0.440	
WORKING MEMORY	9.5017	10.4662	25.9941	0.908	0.372	
COGNITIVE CONTROL	-4.8513	10.5068	25.9938	-0.462	0.648	
TIME:ITEM TYPE	0.6705	4.1181	5626.93	0.163	0.871	

REML criterion at convergence 67216.7

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT INTERCEPT is set to pre-test, baseline items, mean working memory, mean retention, and mean cognitive control. WORKING MEMORY, COGNITIVE CONTROL, and RETENTION have been normalized using Z-transformation. Formula: rt ~ time * itemtype + scale(retention) + scale(wm) + scale(cc) + (1|item) + (1|participant).

3.5 Discussion

The present study looked at the very initial stages of vocabulary learning in second language acquisition. As noted in the introduction, it has been suggested that in the early stages of SLA, formal effects of CLI dominate. On the other hand, at later stages of learning, meaning-based effects of CLI dominate (Bardel, 2015; Ringbom, 2007). It was hypothesized that the findings might be dependent on whether one focuses on overt (as opposed to covert) effects. The present dissertation focuses on meaning-based effects over the course of acquisition. For this reason, both accuracy and processing (to include covert effects) were included in the design of the present study. Both accuracy and processing were included for forward CLI while only processing was included for reverse CLI as the learners were not assumed to make overt errors in their L1 as a result of approximately one hour of L2 instruction.

No significant overt meaning-based CLI was found for *forward CLI* operationalized as accuracy in the L2. Significant effects of meaning-based CLI in processing were found, which highlights the importance of including unconscious lexical activity when estimating CLI (c.f., Jarema and Libben, 2007). It is clear that the L1 affects the L2 in the MML at the early stages of SLA. The significant meaning-based effects in L2 processing support the accounts of a shared conceptual system for languages (c.f., Hall & Ecke, 2003; Jiang, 2000; Kroll & Stewart, 1994; Pavlenko, 2009) at least at the initial stages. The findings also align with Stage A1 in the PM.

Previous studies (Degani & Tokowicz, 2010b; Eddington & Tokowicz, 2013) found that learning items with translation ambiguity in the L2 is more difficult than learning items without translation ambiguity. In the present study, there were some effects of increased difficulty, but all-in-all, the participants managed to learn the items with translation ambiguity to a very high *accuracy*.

The learning experience in the present study diverts from a more naturalistic learning experience in that in the latter case the interval between exposure to specific items is typically longer. Degani, Tseng, and Tokowicz (2014) found that learning the two meanings of translation ambiguous words in close proximity mitigates their learning disadvantage. Given the near back-to-back presentations of the different form-meaning combinations for the items with translation ambiguity, it was hypothesized that it might be possible for Stage A3 in the PM to take place already before Stage A2.

Indications of initial noticing of differences between the L2 form and the host representation (see Stage A3 in the PM) were found both during learning and assessment. In the case of learning, the first presentation of the second meaning had an estimated increase of 194ms in ToTs ($p = 0.71$). In the case of assessment, the first presentation of the second meaning had an estimated increase of 674ms in ToTs ($p = 0.10$). At the chosen significance level ($\alpha \leq 0.05$), the statistical models do not support the hypothesis that the effects of Stage A3 in the PM could appear already before the effects of Stage A2. In the case of the latter analysis of the effects in assessment, given the magnitude of the effect, it is possible that the lack of significance is due to small sample size. Assuming that the participants had detected the difference between the L2 form and the host representation, the prediction was that the participants would then either rehearse activation patterns, as suggested by the PM, or attempt to deal with the ambiguity using introspection, i.e., the use of conscious rules (as suggested by Jiang, 2002). In the case of the former, the difference between the two item types would be expected to disappear, while in the latter the expectation was that the effect would sustain until Stage A2. Accuracy in the items with translation ambiguity surpassed that of the unambiguous items. However, in processing cost, the effect sustained all the way until after Stage A2 (and consolidation).

Another direction of interest in the present study was *reverse CLI*. The point of departure was that should reverse CLI from later acquired languages manifest itself in the L1 in advanced speakers at an overt level, then it should be possible that reverse CLI could be observed in processing even in early stages of vocabulary learning. However, it was hypothesized that the effects would likely be so subtle that they would be difficult, if not impossible, to measure. Very small effects (see Figure 19) were found in the processing of the L1 equivalents of the two translations in Kontu v1. Corrected for generic learning effects, based on the baseline items, we can note that there is an initial increase in RTs between the two translation equivalents in the L1. This represents increased processing difficulty (possibly competition, see Section 2.3.2). After the consolidation period, there is a facilitation effect in the L1 for the two translation equivalents in comparison with the baseline items. However, neither of these effects is significant. The former effect would align with (temporary) cross-language competition – likely caused by the intense immersion in the TL Kontu v1 immediately preceding the post-test – at the initial stage before consolidation.

Finally, *aptitude* was of interest as a predictor of learning outcomes. It was hypothesized that both CC and WM would be relevant predictors for learning outcomes, as they have been found to predict language learning capacity (Ellis & Sinclair, 1996). With respect to CLI from L1 to the L2, the effect of CC and WM was hypothesized to be different. Increased WM, as a type of a short term memory, could aid leaners in the assessment modules and possibly in the post-test. Based on Green's (1998) IC account, increased CC capacity should correlate with less CLI from the L1 to the L2. For L2 to L1 CLI in MML at the initial stages, there is little previous research to build upon. The results showed that the effect of aptitude seems to only relate to processing cost rather than attained accuracy. As pointed out, the participants acquired both items with translation ambiguity and without translation ambiguity with little difficulty. Items with translation ambiguity had higher processing cost and this increase was modulated by available WM resources. No statistically significant effects for CC were found.

3.6 Conclusion

The aim of the present study was to investigate forward and reverse CLI and the effects of aptitude at the very initial stages of second language acquisition. For this purpose, a pseudolanguage 'Kontu v1' was created. The study was experimental in nature. The participants were *ab initio* monolingual L1 speakers of English.

No statistically significant overt meaning-based CLI in accuracy was found, but statistically significant effects were found in processing (i.e., longer time on task) – highlighting the importance of including unconscious lexical activity when estimating CLI. Controlling for general learning effects, reverse CLI did present itself in the data but in quantities that fall well within the margin of error.

Finally, previous research had found that words with translation ambiguity are harder to acquire. This finding was not replicated in the present study with respect to accuracy. Furthermore, the results indicate that the participants were capable of acquiring Kontu v1 irrespective of aptitude, but that aptitude was associated with how effective the participants were at resolving ambiguity. Also, only WM was a significant predictor of processing cost. It can be hypothesized that for cognitive control to play a role for the processing of translation ambiguity, further entrenchment and consolidation has to take place. The participants were also novel second language learners, meaning that they would not have had to acquire skills to regulate CLA. It will be interesting to see whether these patterns are similar in third language acquisition, which will be addressed in Study 2 (Chapter 4) below.

4 Cross-Linguistic Influence in Early L3 Word Learning

4.1 Introduction

Study 1 (Chapter 3 in this dissertation) looked at *cross-linguistic influence* (CLI) in the *multilingual mental lexicon* (MML) at the very initial stages of *second language acquisition* (SLA). Meaning-based CLI was found in *processing* (time spent on task) but not in *accuracy* (proportion between correct and incorrect responses). Both forward (L1 to L2) and reverse (L2 to L1) CLI were of interest. Statistically significant effects were found for forward CLI but not for reverse CLI. There were, however, patterns that suggested initial, temporary, lexical *competition* effects, and later (post-consolidation) *facilitation* effects in the processing of the L1. That is, controlled for a learning effect based on performance in the baseline items, the participants took longer to respond to items with translation ambiguity between the L1 English and the L2 Kontu v1 after input in the L2. On the other hand, after a one-night consolidation period, this effect was reversed. However, the quantity of these fell well within the margin of error. One interesting question is whether, at the initial stages, the L2 is less resistant than the L1 for reverse CLI. The present study is essentially a replication of Study 1 with speakers that, at the initial state of the acquisition, already have two proficient source languages (L1 Swedish and L2 English) to build upon and to regulate.

Forward CLI refers to the influence of the previously acquired language(s) observed in the later acquired language(s). This means that when investigating forward CLI, the *source languages* (SLs) in the present study are the L1 Swedish and the L2 English, and the *target language* (TL) is the L3 Kontu v2 which is a pseudolanguage. *Reverse CLI* refers to the influence of the later acquired languages observed in the chronologically previously acquired languages. That means that when investigating reverse CLI, the SLs in the present study are the L2 English and the L3 Kontu v2, and the TL is the L1 Swedish. Out of the six potential directions of CLI between the three languages, the following three directions were investigated in the present study: 1) forward CLI from the L1 and the L2 (jointly) to the L3 Kontu v2, 2) reverse CLI from the L3 Kontu v2 to the L2 English, and 3) reverse CLI from the L3 Kontu v2 to the L1 Swedish. That is, the stimuli has purposefully been designed so that the difference between the L1 and the L2 compared to the L3 is the same (to the extent that it is possible considering interlanguage effects). This makes an analysis of the two latter directions comparable, whilst removing the possibility to analyze whether forward CLI takes primarily place from the L1 or the L2.[42]

As with Study 1, the *Parasitic Model* (PM; Hall & Ecke, 2003) would predict immediate effects of forward CLI since the new words in the L3 are presumed to be connected to L1 and/or L2 host representations (see Section 2.2.2). The host representation is the L1 or the L2 word that is determined to be the closest match – either based on form, if possible, or meaning, in which case the L3 form connects to the conceptual representation via the L1 and/or the L2 frame. Later, differences between the three languages are detected at the item level, presumably mitigating CLI. Furthermore, the learners will have to acquire skills to regulate cross-language activation (CLA; see section 2.1 for accounts of *priming*, *spreading activation,* and *non-selective access*).

In terms of methodology, many aspects of Study 1 were integrated into Study 2 to allow for comparison of the results. However, a new version of the pseudolanguage (henceforth referred to as "Kontu v2"), was created due to the addition of Swedish in the experiment. The addition of Swedish required that the items in the pseudolanguage had to be controlled and adapted based on Swedish in addition to English. Two items were replaced, and sixteen items were added (the changes are outlined in Section 4.3.2.1 below). Kontu v2 consists of 48 words with 52 meanings and, like Kontu v1, is phonotactically adherent with Finnish. The discrepancy in the number of forms and meanings is due to four forms corresponding to two translation equivalents in the L1 and the L2, representing *translation ambiguity*.

[42] Whether forward CLI takes place from the L1 or the L2 in L3 lexical acquisition and processing has been much more widely researched than the distinction of the latter two directions in reverse CLI, for which reason this choice was made. For previous investigations on the primary forward supplier in L3 acquisition, Bardel (2015), Jarvis & Pavlenko (2008), and (Ringbom (2007) provide extensive overviews.

The acquisition of Kontu v2 words was tracked throughout the learning, as was forward and reverse CLI. Furthermore, the effects of *aptitude* and *psychotypology* were an object of interest in the study.

4.2 Aim and predictions

The main purpose of the present study was to replicate Study 1 with multilingual learners who had already acquired a second language and for whom Kontu v2 is the L3. Given this aim, the following research questions were formulated:

RQ1 If present, is cross-linguistic influence (CLI) at the initial stage in the multilingual mental lexicon (MML) unidirectional (only forward) or multidirectional (both forward and reverse)?

RQ2 To what extent is CLI in the MML affected by aptitude?

RQ3 To what extent is CLI in the MML affected by psychotypology?

The first research question refers to the directionality of CLI at the very initial stage. Three potential directions were tested: 1) forward influence from the previously acquired languages L1 Swedish and L2 English to the L3 Kontu v2, 2) reverse influence from the L3 to the L1, and 3) reverse influence from the L3 to the L2. Increasing proficiency at the initial stages was operationalized as increased knowledge in the L3 Kontu v2. Despite increasing proficiency in the L3, the assumption was that at this stage of learning, CLI will likely be disproportional in that it takes place quantitatively more from the previously acquired languages to the L3 than vice versa. The second research question relates to the effect of aptitude, which is one of the modulating factors that are of interest in this dissertation. Aptitude in the present study is operationalized as *working memory* (WM) and *cognitive control* (CC). The third research question refers to the role of psychotypology, which in the present study is operationalized both as *assumed similarity* and *perceived similarity*. Both are measured and treated as factors with between-participant variation in the sample population. The former is measured prior to and the latter after learning Kontu v2.

Explicit predictions about the acquisition of L3 words at the very initial stage are made by the *Parasitic Model* (PM, presented in Section 2.2.2). As shown in Table 15, four processes are hypothesized to take place at the first stage of learning a particular word in the L3. In Stage A1, a learner encounters a new word in the L3, and the learner makes use of the *short term memory* (STM) creating a connection between the L3 form to what is perceived to be the closest match in any of the learner's languages. This is done either based on form, if possible, or meaning (Stage A4). Then, in Stage A2, the connection between the L3 form and its *host representation* is established in the *long term memory* (LTM). In Stage A3, differences between languages are detected. For successful performance in resolving translation ambiguity, without the use of conscious rules, Stage 3 would be required since the connected L1/L2 representation is not compliant with conventional use of the word in the L3.

Table 15

First stage of the Parasitic Model of Lexical Development (Hall & Ecke, 2003)

A1	A2	A3	A4
The L3 word form is registered in STM and the closest matches in L3, L2, or L1 are activated based on salient form attributes.	The L3 form is connected to a host representation and is established in LTM in distributed fashion.[43]	Differences between L3 form and host representation are detected, new patterns are rehearsed and the representation is revised with respect to the attributes that distinguish it from the host and/or other consolidated neighbors.[44]	If no matching representation is activated sufficiently, the L3 form is connected to the frame of the nearest conceptual equivalent.

The PM would predict that at least Stage A1 takes place during the learning phase in the present study. This was the case in Study 1. The additional variable in the present study is that it is now possible for the learner to attach the form to a representation in one of the two pre-existing languages. While the PM opens for the possibility that the learners can attach the new L3 form to a pre-existing L3 host, given that the learners are at the initial state in the L3 at the onset of their participation, this possibility is theoretical at best. The stimuli in the present study have been designed so that whether the participant chooses a L1 or a L2 representation at this stage, the resulting L3 form-meaning mapping is as similar

[43] The *host representation* is normally the most highly activated related L1 form, where some threshold level of similarity between the items is met. *Distributed fashion* refers to the activation of same nodes in the network as the host form.

[44] Not always achieved, leading to fossilization of the interlanguage configuration (cf. Jiang, 2000).

as possible. The processing of meaning in the L3 is expected to take place via either the L1 Swedish or the L2 English. For this reason, we would expect forward CLI to take place from the very onset, which was the case in Study 1.

One of the questions in Study 1 was whether the intensive teaching format (short interval between exposure to items) is able to produce effects from Stage A3 already during the learning process despite Stage A2 having not yet taken place. This would facilitate the learning of the items with translation ambiguity. The results were inconclusive. The question for Study 2 is whether the L3 learners, who due to already having acquired two languages, might be better able to deal with translation ambiguity from the onset. In previous research, bilinguals have been found to be better word learners (Kaushanskaya & Marian, 2009; Papagno & Vallar, 1995; Van Hell & Mahn, 1997). One hypothesis is that this is caused by greater sensitivity to semantic information during learning (Kaushanskaya & Rechtzigel, 2012). Ringbom (2007) hypothesized that a bilingual advantage in word learning is related to better metalinguistic awareness. The participants in Study 1 managed to acquire the Kontu v1 items, with and without translation ambiguity, to high accuracy so from that perspective the interesting question is whether bilinguals fare better in processing. Hence, as with Study 1, the assumption is that CLI *will* take place and that processing in the L3 *is* affected by the L1 and/or L2-derived lexical-conceptual representations. This means, that the primary question for forward CLI is not so much *whether* there is forward CLI, but rather *how* it operates from the perspectives of accuracy and processing at the very initial stages – as well as how the bilingual learners in the present study differ from the *ab initio* monolingual learners in Study 1.

In *reverse CLI*, only processing is investigated in the present study. In Section 2.3.2, three possible explanations for negative reverse CLI, and one potential explanation for positive reverse CLI were provided. The presented negative explanations were: 1) increase in activation thresholds in line with the *Activation Threshold Hypothesis* ATH (Paradis, M., 2007), 2) increased need for regulation due to increased amount of competitors in line with the *Inhibitory Control* (IC) model (Green, 1998), and 3) retrieval of Ln words impairing their L1 equivalents (Levy, McVeigh, Marful & Anderson 2007). The presented explanation for positive reverse CLI was an *indirect frequency effect* where activation of Ln words facilitates the processing of their L1 translation equivalents (Higby, Donnelly, Yoon & Obler, 2020). The first explanation of negative effects relates to attrition in the form of decay. No extended periods without use of the L1 or the L2 are associated with either Study 1 or Study 2, and thus such effects are not expected even theoretically. The second and third accounts of negative effects, as well as the account of positive effects, are caused by cross-language activation (CLA). Thus, these can, potentially, explain even early reverse CLI. However, none of these make explicit predictions about the effects of translation ambiguity at the very initial stages. Given the lack of specific predictions, as with Study 1, the question in the present study is *whether* any reverse CLI can be observed at the initial stages.

Additionally, the effect of *aptitude* (operationalized as WM and CC) is of interest in the present study. Better WM would be expected to have a facilitative effect in learning whereas better CC should aid the learners in regulating the L1 and the L2 (and presumably the L3, should it require any regulation) potentially leading to less CLA, and subsequently, less CLI. In Study 1, the participants were capable of acquiring Kontu v1 successfully irrespective of aptitude. There was, however, an effect of WM for the participants' ability to resolve translation ambiguity. No effects for CC were found. It was hypothesized that this might be related to the participants being both overall novel language learners as well as initial state learners of Kontu v1, meaning that they would not have had to previously acquire skills to regulate CLA. The participants in the present study have two pre-existing languages. From this perspective, we would expect them to be more dependent on CC than the participants in Study 1.

Finally, the effect of *psychotypology* is of interest in the present study. Psychotypogy (perceived cross-linguistic similarity, see Section 2.1.2), is treated both as *assumed similarity* and *perceived similarity* (c.f., Kaivapalu, 2004, Kaivapalu & Muikku-Werner, 2010). The former, in the present study, is tested before exposure to Kontu v2 and the latter after exposure to Kontu v2. The former tells us about the learners' pre-existing, conscious, conceptions that are measured explicitly, and that are hypothesized to influence their learning. The latter tells us about from which of the SLs (L1 and L2) in forward CLI the participants are more *willing to accept* influence from after exposure to Kontu v2, and is tested implicitly. Ringbom (2007) hypothesized that perceived similarity (as opposed to objective similarity) in language learners is not necessarily symmetrical with respect to direction. In line with this possibility, both assumed and perceived similarity are used as predictors in the statistical models for forward and reverse CLI.

4.3 Method

4.3.1 Participants

The participants were Swedish (L1) university students with high proficiency in English (L2, ≥ CEFR B2) and limited knowledge (≤ CEFR A2) of other languages. Thirty participants took part in the study. The participants filled in the same two-page language background questionnaire as in Study 1 (see Appendix A) and the responses were used as exclusion criteria.

The participants' mean age was 24.6 (Min-Max: 19-36, *SD* = 3.61), and 47% were female. All had grown up in Sweden and had not spent extended periods outside of Swedish- or English-speaking countries. The participants were studying a variety of different subjects at Lund University at the time of testing. Most participants were undergraduate students. The participants self-evaluated their language learning aptitude at 4.81 (Min-Max: 3-6, *SD* = 0.75), language learning attitude at 5.75 (Min-Max: 2-7, *SD* = 1.29), and ability for cross-cultural communication at 5.53 (Min-Max: 4-7, *SD* = 0.85) on a seven-point Likert-scale.

4.3.2 Instruments

4.3.2.1 Adjustments to instruments from Study 1

Most of the instruments used in Study 2 are identical to those in Study 1. These include the background questionnaire, the learning and assessment tasks, as well as CC and WM tasks (see Section 3.3.2). The stimulus set was slightly altered due to the addition of Swedish: the items had to be comparable across the L1 and the L2. Since there was no consolidation period, the sleep quality questionnaire was omitted (see Section 4.5 for a lengthy discussion of the reason for this). Two new tasks were included to estimate psychotypology which are presented below. Furthermore, the priming tasks were translated to Swedish and administered in both the L1 Swedish and the L2 English. All-in-all, the addition of Swedish and psychotypology instruments, and due to bilinguals being slower in lexical access than monolinguals (Portocarrero, Burright & Donovick, 2007; Sandoval, Gollan, Ferreira & Salmon, 2010), the length of the experiment increased by approximately 50%.

The new version of the pseudolanguage, Kontu v2, consists of 48 word forms taken from Finnish. The 32 word forms from Kontu v1 are identical in Kontu v2. Twenty eight of these words had a clear dominating translation equivalent in the L1 and the L2. In addition, four word forms had translation ambiguity in the sense that they had two translation equivalents both in the L1 and the L2. This gives us 36 meanings: 28 words with one unitary meaning (28 x 1), and four with two meanings (4 x 2). All translation equivalents were primary referents of nouns that could be easily depicted using pictures. All form-meaning mappings for these 32 words are identical with Study 1 except for two meanings that had to be changed due to Swedish operating differently from English. Given that the additional 16 forms with surface overlap were not taught to the participants, the amount of items (as well as their forms) taught to the learners in Study 1 and Study 2 is identical.

The additional 16 word forms in Kontu v2 (presented in Table 16 under I-CLSA below) were used for a psychotypology task and were not taught to the participants. In line with the PMs prediction of form attributes predicting initial choice of which L1/L2 representation is activated, these 16 items were designed to have surface overlap with either the L1 or the L2. Ten Swedish-English bilinguals confirmed the transparency of their meanings and 10 L1 Finnish speakers that their form was acceptable based on Finnish phonotactics (i.e., that they were acceptable as potential loan words in Finnish).

Finally, the set of 16 English pseudowords from Study 1 was also used and an additional set of 16 pseudowords in Swedish was created. The pseudowords were matched in both syllable quantity and length in letters to the Swedish forms for the items in the experimental set. They were adjusted from the original words by substituting consonants and vowels. Ten L1 Swedish participants confirmed their status as non-words in Swedish. Furthermore, a set of eight non-words in Kontu v2 was created representing extreme violations of Finnish / Kontu v2 phonotactics (presented in Table 16 under I-CLSA below). Their non-acceptability in Finnish was confirmed by 10 L1 Finnish speakers.

A full list of form-meaning mappings in Kontu v2 is provided in Appendix F and a matrix of primes and targets is provided in Appendix G.

4.3.2.2 *Explicit Cross-Linguistic Similarity Assessment (E-CLSA)*

Psychotypology can be *assumed* or *perceived* (Kaivapalu & Muikku-Werner, 2010), and it can be measured using *explicit* and *implicit* means. Assumed psychotypology refers to a learner's preconceptions about similarity between two languages, while perceived psychotypology relates to a learner's evaluations of similarity based on exposure. When psychotypology is measured explicitly, conscious evaluations are elicited from language learners, while in implicit tasks, the learners' estimations of similarity are probed using non-conscious means.

Few standardized research instruments for the assessment of psychotypology exist in the literature. The Explicit Cross-Linguistic Similarity Assessment task (E-CLSA) was developed to estimate language learner's conscious evaluations of language similarity. It can be used both to measure assumed and perceived psychotypology depending on whether it is administered prior to the onset or after some exposure to the language.

Thirty six area-specific statements were used covering the following areas of language similarity: global similarity, grammar, idioms, phonology and pronunciation, spelling, and vocabulary. This allows for both an analysis of psychotypology as an integral, language-as-a-whole, unit as well as independently for specific subcomponents of language knowledge. The statements are based on Neuser (2017), Haghverdi et al. (2012), Suhonen (2015), and Sayehli (2013). A full list of the statements is available in Appendix H. For the purposes of the present study, given the small number of participants, a compound score of psychotypology was used in the analysis. In the compound score, responses from each statement are given the same amount of weight.

In the E-CLSA, the participants were asked to disagree or agree with the statements using a seven-point Likert-scale.

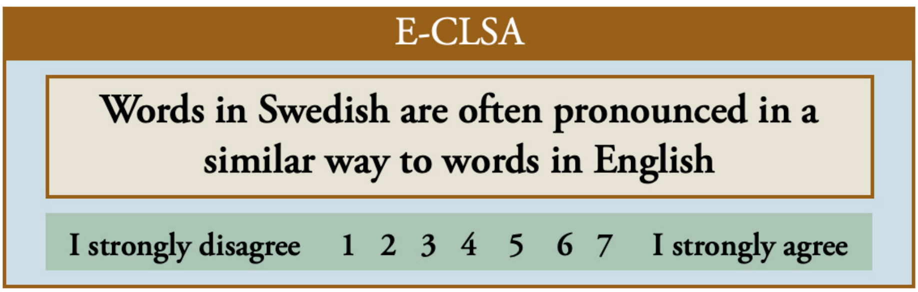


Figure 20. Stimulus presentation in E-CLSA

The E-CLSA was delivered using Superlab 5 software and a Cedrus RB-740 response pad. The participants were presented with the statements about similarity between the three languages like in Figure 20 above and used the response pad to state their answer on the seven-point scale, visually marked on the buttons. The order of the statements was randomized for each individual participant.

4.3.2.3 Implicit Cross-Linguistic Similarity Assessment (I-CLSA)

As suggested in Section 4.3.2.2, psychotypology can be *assumed* or *perceived*, and it can be measured using *explicit* and *implicit* means. In addition to evaluating the participants' explicit assumptions of similarity between Swedish, English, and Finnish, the participants' implicit perceptions of similarity between Swedish, English, and Kontu v2 were tested after exposure to Kontu v2. Three sets of Kontu words were created. The first set of eight items did not adhere to Finnish phonotactics (with extreme violations) and represents non-words in Kontu.[45] Sixteen items in Kontu were created so that eight of them were noticeably derived from English and eight from Swedish. These were adjusted to follow Finnish phonotactics using typical conformational strategies for loan words in Finnish described in Laalo (1990). They were evaluated by ten native Finnish speakers to be acceptable in Finnish. Ten bilingual Swedish-English speakers confirmed that their meaning was transparent despite adjustment for Finnish phonotactics.

In the Implicit Cross-Linguistic Similarity Assessment Task (I-CLSA), the participants were presented with familiar words from the learning component in Kontu v2 as well as unfamiliar words belonging to three categories: 1) nonwords that do not follow Finnish phonotactics, 2) words derived from Swedish adjusted to follow Finnish phonotactics, and 3) words derived from Swedish adjusted to follow Finnish phonotactics. The purpose of the first stimuli was to mask the manipulated variation so that the participants would assume that the comparison was

[45] Since Kontu v2 is a pseudolanguage, technically all words in Kontu v2 are non-words, but a distinction is made here between items that are *supposed to* represent "real" words in Kontu v2 and items that represent "non-words" in Kontu. The latter category is distinguished from the former in that these items do not follow the phonotactics of Kontu v2 (identical phototactics with Finnish). The violations are so extreme (see Table 16) that the participants, who for social reasons, have some pre-existing knowledge of Finnish phonotactics, are expected to be able to distinguish between the item types. Even in the case they cannot do so based on their pre-existing knowledge, based on the results from Gullberg et al. (2010), where only seven minutes of exposure was enough to acquire some basic phonotactic knowledge, the approximately one-hour of exposure to Kontu should be ample enough for the acquisition of basic Kontu v2 phonotactics.

between Kontu v2 words and non-words. The stimuli in the I-CLSA is presented in Table 16 below.

Table 16

Kontu v2 stimuli in the implicit cross-linguistic similarity assessment task

Non-Words	Swedish-based	English-based
paze	sitrooni	lemoni
vlasy	leijooni	lioni
srdce	putiikki	marketti
jazyk	haavetti	oseani
pouzivat	appelsiini	orangiini
slozity	löökki	onioni
oblicej	pumppaani	pumpkini
vpred	kostyymi	suitti

Note. The non-words violate Finnish phonotactics. The Swedish- and English-based words are manipulations from the respective language to Finnish phonotactics and were evaluated by native speakers of the respective languages as being legal in the given language. All the above forms were novel to the participants.

The task was disguised as a proficiency test in Kontu v2. The participants responded to the lexical stimulus in Kontu v2 using a button box. Visualization of the stimulus presentation is provided in Figure 21 below.

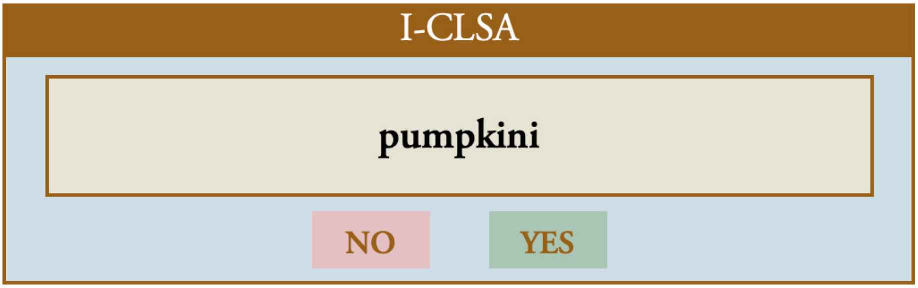


Figure 21. Stimulus presentation in I-CLSA

The Kontu v2 word 'pumpkini' is derived from the English word 'pumpkin,' presented in the nominative case, and adjusted to Finnish phonotactics using typical conformational strategies for loan words in Finnish described in Laalo (1990).

The order of presentation of the items was randomized for each individual participant. The stimulus was always displayed centered on the screen using SuperLab 5. Auditory feedback was provided with a positive answer for the non-words resulting in a buzzer sound. The participants responded to the stimuli using two color-coded (red, green) buttons using a Cedrus RB-740 response pad.

4.3.3 Procedure

First, the participants were informed about the purpose of the study and were allowed to ask questions and asked to sign a consent form. After it was mutually agreed that data collection could commence, the participants were asked to fill in the background questionnaire. While the participants had been pre-screened, one participant had to be excluded due to their responses in the background questionnaire. After the background questionnaire, the participants took the E-CLSA task.

The experiment consisted of five main sections (visualized in Figure 22 below).

BACKGROUND ENGLISH	Welcome Consent for participation[46] Background questionnaire E-CLSA (psychotypology)
PRE-TEST SWEDISH & ENGLISH	Priming task in L1 Swedish (functions as a baseline) Priming task in L2 English (functions as a baseline)
LEARNING KONTU	Learning Kontu v1 (four learning blocks and four assessment blocks)
POST-TEST SWEDISH & ENGLISH	Priming task in L1 Swedish (identical to pre-test) Priming task in L2 English (identical to pre-test)
INDIVIDUAL DIFFERENCES ENGLISH	I-CLSA (psychotypology) Flanker (cognitive control) Working memory (*n*-back) Debriefing

Figure 22. Procedure in Study 2

[46] The participants were provided with a privacy policy and data management description. In accordance with GDPR, the European Union General Data Protection Regulation (2016/679), the participants consented in writing to the following: 1) that their personal information will be stored separately from research data, 2) that personal information will be stored in relation to their compensation in accordance with the Swedish accounting legislation (BFL 1999:1078) by Lund University centrally, 3) the collected but anonymized research data will be publicly available and may be used for further research purposes without additional consent from the participant, 4) and that it is possible to retract one's participation up until the point of publication. The participants were compensated with movie tickets for their participation. None of the conditions set forth by The Swedish Research Council for projects to be reviewed by an ethics board apply for the given study.

These were the background section, the pre-test, the learning of Kontu v2, the post-test, and the individual differences measures.[47] Throughout the procedure, two types of breaks were given. Between major sections, the participants were instructed to take breaks involving physical movement. The pre-test, the learning component, the post-test, and the aptitude tasks also included assigned microbreaks where the participants were instructed to rest their eyes for a moment to ensure that they could concentrate efficiently. The pre-test functions as a baseline to evaluate change in the priming effect of the experimental items with translation ambiguity in relation to the baseline items. That is, there are two 'baselines' against which participant performance is being compared with: one for time and one for items. The presentation order in the priming task was randomized for each participant. All priming tasks were introduced as decision-making speed tasks. After the learning module, which was identical with Study 1, the participants took the priming tasks in the form of post-test. These were identical with the ones in the pre-test except that item randomization was done again for each participant.

In the last section, the participants took the I-CLSA, WM, and CC tasks. Finally, debriefing was performed. Almost all participants reported some awareness of the translation ambiguity and some were aware of some form of presentation of stimulus between the hashes and the targets. No participant reported having been able to (consciously) read the primes. The participants were actively informed about the primes, the purpose of the priming task, and that Kontu v2 is not the same as (some simplified form of) Finnish at this point.[48] All-in-all, the data collection took approximately 90 minutes for most participants, although for a few participants it took in the excess of 120 minutes.

---

[47] In Study 1, there was a delayed post-test after a one night consolidation period and the aptitude components were administered as a part of that section. In the present study, the aptitude measures (WM and CC) were administered immediately after the other experimental tasks.

[48] *Psychotypology* in the present study was operationalized both as *assumed* and *perceived* similarity. Pre-existing conceptions about the Finnish language were used to probe assumed similarity. Thus, for the purposes of researching psychotypology, the participants were left under the impression that Kontu v2 is some form of "simplified" Finnish (without grammar) for the duration of the experiment. However, the participants were explicitly informed in the debriefing at the end of the data collection session that the form-meaning mappings they had acquired do not reflect those in Finnish.

4.3.4 Data analysis, transformations, and structure

There are four outcome variables in Study 2: 1) time on task (ToT) in the learning sequences, 2) ToTs in assessment, 3) accuracy in assessment, and 4) response times (RT) in the priming task. The first three represent forward CLI and the last reverse CLI. A detailed description of distribution of the data in Study 2, before and after any trimming, is provided in Appendix I. All statistical analysis was performed in R version 3.6.0 (R Core Team, 2018) using R Studio version 1.2.1335 as the graphical user interface. Data exploration was done using Jamovi version 1.0.8.0 (The Jamovi Project, 2019).

The ToT data for learning sequences and assessment (variables 1 and 2 above) contained in total 21,276 observations (with 324 missing values). Lower-spectrum trimming was set to 100ms (n = 363) which represented 1.68% of the data. This left 20,913 observations. The two conditions were almost equally affected with 96.85% data left in the baseline condition and 96.73% data left in the condition with translation ambiguity. No upper spectrum trimming was done since the primary purpose of the learning segments was to maximize the participants' learning of the items and as such time limits were not enforced.

The learning outcome data was collected on four occasions spaced after every four instances of each item (of a total of 16). This data consists of binary accuracy data from the learning outcome test. Chance in the test was set at one in four. Overall, accuracy in the test, like in Study 1, was extremely high and points to the participants having taken the learning task very seriously. No trimming or transformations were performed on the accuracy data.

The response time data set for the priming tasks (Swedish and English) contained in total 13,920 observations. Some 252 (1.81%) values were missing due to technical issues. The participants in Study 2 were overall slower in the priming task than the participants in Study 1, which is expected given slower lexical access in bilinguals as well as that the participants did the task both in their L1 and L2. For this reason, a higher upper trim point was chosen. Response times of <200ms and >1,400ms were removed (n = 98) which represented 0.70% of the data. Distractor items were most affected (1.4%). The items used to analyze reverse CLI were almost equally affected at 0.67% for baseline and 0.52% for items with translation ambiguity.

The aptitude data in consists of self-evaluations, WM scores, and CC scores. Upper spectrum trimming was not necessary for the RT data as both the WM and CC tasks timed out at 2,000ms (n = 185). RTs <100ms were removed from both

WM and CC scores (n = 211). Both WM and CC data were normalized using z-transformation for the linear mixed effect models.

The psychotypology data consists of two sets of data: 1) 1,080 measures of perceived language similarity between Swedish, English, and Finnish on a seven-point Likert scale, and 2) 1,680 RT measures on familiar items, Kontu v2 non-words, and novel lexical items in Kontu v2 based on the two potential source languages. These were computed into participant-specific explicit and implicit psychotypology scores. The explicit measure was computed by first transforming negative and positive statement responses to a scale where higher always means more similar and then deducting the mean of Swedish-Finnish similarity evaluations (1-7) from the mean of English-Finnish similarity evaluations (1-7), resulting in a score of -6 to 6 where 0 is a perfectly balanced score. For the implicit measure, in line with the priming data, RTs <200ms and >1,400ms were removed representing 2.56% of the data. The final implicit measure was computed by deducting the mean of the Swedish-based items from the mean of the English-based items resulting in a score of -6 to 6 where 0 is a perfectly balanced score.

4.4 Results

4.4.1 Forward CLI in L3 Kontu

Forward CLI in the present study refers to the effect of L1 Swedish and L2 English on L3 Kontu v2. There are three measures that can capture such effect: 1) processing cost (ToTs in learning across item types), 2) processing cost (ToTs in assessment across item types), and 3) accuracy (correct responses in assessment across item types).

As to the first measure of forward CLI in L3 Kontu, ToTs in learning sequences were tracked across item types over the course of learning. In line with Study 1, ToTs reduce over time. Figure 23 below presents the ToTs for the two types of stimuli in the learning task throughout their sixteen occurrences.

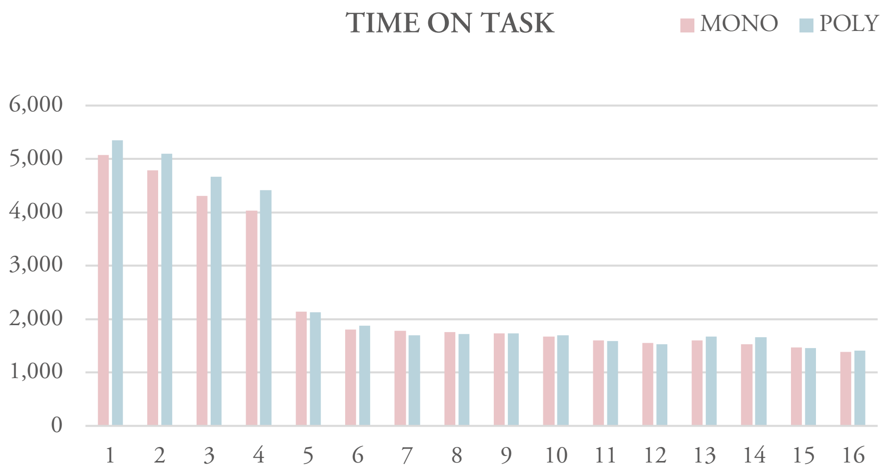


Figure 23. Time spent on task (ms) in learning sequences

The numbers 1-16 refer number of times an item has been presented in Kontu v2. MONO refers to items without translation ambiguity and POLY to items with translation ambiguity. The values represent means across items and participants. Table I6 in Appendix I.

In Study 1, ToTs for the two item types differed from each other on the first instance as well as the first instance after the first assessment module. It was hypothesized that this might, at least partially, be explained by restructuring (Stage A3 in the PM) and forward CLI from the L1 English. When controlled for random effects from item and participant, the estimated effect was 195ms but it was not significant in the statistical model. In the present study, we see that the effect of increased time on task for items with translation ambiguity sustains all the way until the first assessment module. After the first assessment module, there does not seem to be any major differences between the two item types. In Study 1, there was an estimated effect of 125ms between the item types in time spent on task during learning but this difference was not statistically significant.

A linear mixed effect model (LME) was performed to test whether the two item types differed in time spent on task in Study 2. Random intercepts were included for PARTICIPANT and ITEM. The main (fixed) factors TIME and ITEM TYPE were added as interactions. An effect for ITEM TYPE entails CLI and an interaction between TIME and ITEM TYPE entails that this effect changes over the time course of learning. WM, CC, I-CLSA, and E-CLSA were added as co-variates in line with the second and third research questions. The model is presented in Table 17 below.

The estimated effect for ITEM TYPE (262ms) is approximately two times as large as in Study 1 and it is also significant ($p = 0.02$). The interaction between TIME and ITEM TYPE is also approximately two times larger in magnitude compared to Study 1 and is also significant ($p = < 0.01$), which entails that the participants are either more affected by the translation ambiguity between the TL and the SLs (the TL items have translation ambiguity towards both L1 and L2) or their better metalinguistic awareness of translation ambiguity between languages results in the learners actively paying more attention to these items in learning.

Table 17

Forward cross-linguistic influence in processing cost over time in learning

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.0567	-0.4799	-0.0971	0.2902	22.7315

Random effects

GROUPS	NAME	VARIANCE	SD
ITEM	INTERCEPT	43074	207.5
PARTICIPANT	INTERCEPT	595616	771.8
		NUMBER OF OBS: 16695, PARTICIPANT, 30	

Fixed effects

	ESTIMATE	SE	df	t	*p*	
INTERCEPT	4351.3461	193.7909	28.7770	22.454	< 0.001	***
TIME (INSTANCE)	-226.0852	3.6682	16630.96	-61.63	< 0.001	***
ITEM TYPE	261.7551	112.2877	80.4990	2.331	0.02225	*
WORKING MEMORY	170.7579	155.7013	25.0159	1.097	0.28322	
COGNITIVE CONTR	10.0944	143.5265	25.0002	0.070	0.94449	
E-CLSA PSYCHOTYPO	204.0902	276.2268	25.0072	0.739	0.46688	
I-CLSA PSYCHOTYPO	-0.5171	1.4951	25.0391	-0.346	0.73233	
TIME:ITEM TYPE	-20.6998	7.7620	16628.06	-2.667	0.00766	**

REML criterion at convergence 299800.6

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to first presentation instance, baseline items, mean working memory, mean cognitive control, balanced e-clsa, and balanced i-clsa. WORKING MEMORY and COGNITIVE CONTROL have been normalized using Z-transformation. Time (generic learning effect) is statistically significant. Formula: rt ~ time * itemtype + scale(wm) + scale(cc) + e-clsa + i-clsa + (1|participant) + (1|event).

As to the second measure of forward CLI in L3 Kontu, the processing cost in terms of ToTs in assessment across item types, in Study 1, items with translation ambiguity required more processing effort. This is the case also in the present study (see Table 18 below). The RTs in the learning assessment modules (LOTs) decrease after each module, but the difference between the single- and two-translation items persists until the end.

Table 18

Response times (ms) in assessment modules by item type

TEST	ITEM TYPE	MEAN	*SD*	MEDIAN
1	MONO	4015	2120	3614
	POLY	4188	2124	3654
2	MONO	2730	1358	2375
	POLY	2868	1425	2433
3	MONO	2495	1458	2097
	POLY	2538	1232	2126
4	MONO	2556	1575	2075
	POLY	2702	1904	2079

Note. MONO refers to items without translation ambiguity and POLY to items with translation ambiguity. The items with translation ambiguity require more processing effort across the board.

A LME was performed to assess the response times across item types over time. Random intercepts were added for PARTICIPANT and ITEM. The main (fixed) factors TIME and ITEM TYPE were added as interactions. An effect for ITEM TYPE entails CLI and an interaction between TIME and ITEM TYPE would that this effect changes over the time course of learning. WM, CC, I-CLSA, and E-CLSA were added as co-variates in line with the second and third research questions. The model is presented in Table 19 below. The main effect for ITEM TYPe (299ms) is approximately half and the interaction for TIME and ITEM TYPE approximately ¼ of that in Study 1. Neither is statistically significant.

Table 19

Forward cross-linguistic influence in processing cost over time in assessment

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.2369	-0.5570	-0.1789	0.3404	9.6258

Random effects

GROUPS	NAME	VARIANCE	SD
ITEM	INTERCEPT	544723	738.1
PARTICIPANT	INTERCEPT	349063	590.8

NUMBER OF OBS: 4218, PARTICIPANT, 30; ITEM, 72

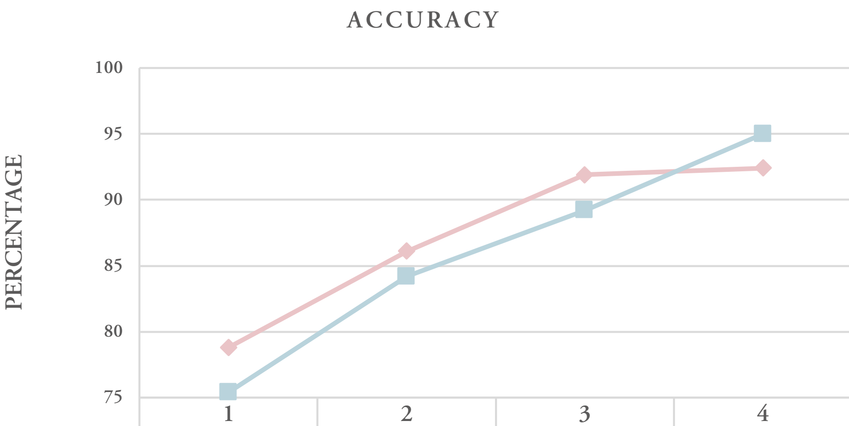
Fixed effects

	ESTIMATE	SE	df	t	*p*	
INTERCEPT	4947.415	204.021	85.953	24.250	< 0.001	***
TIME (BLOCK)	-698.213	34.378	2659.050	-20.310	< 0.001	***
ITEM TYPE	299.265	306.923	193.923	0.975	0.331	
WORKING MEMORY	22.274	121.137	25.041	0.184	0.856	
COGNITIVE CONTROL	168.623	111.640	25.003	1.510	0.143	
E-CLSA PSYCHOTYPOL	86.432	214.840	25.010	0.402	0.691	
I-CLSA PSYCHOTYPOL	0.337	1.162	24.997	0.290	0.774	
TIME:ITEM TYPE	-57.320	72.874	2642.781	-0.787	0.432	

REML criterion at convergence 73671.2

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to first assessment block, baseline items, mean working memory, mean cognitive control, and equal distance from Swedish and English for both E-CLSA and I-CLSA. WORKING MEMORY and COGNITIVE CONTROL have been normalized using Z-transformation. Time (generic learning effect) is statistically significant. Formula: rt ~ time * itemtype + scale(wm) + scale(cc) + e-clsa + i-clsa + (1|participant) + (1|event).

The third measure of forward CLI in L3 Kontu was accuracy in the items in the LOTs. The learning component was primarily designed to ensure that the participants would acquire Kontu v2 words. Accuracy measures from the four LOTs during the learning phase suggest that this is the case. These included four options in the form of pictures, none of which were novel to the learners, and as such chance is at 25%. After the first four instances, the participants performed at 79% and 75% on the single- and two-translation items, respectively. By the fourth and final assessment module, this had increased to 92% and 95%, respectively. The data is presented in Figure 24 below. What we can see is that (like in Study 1) the items with translation ambiguity initially lag behind in accuracy but surpass the items without translation ambiguity in the end.



	1	2	3	4
mono	78.8	86.1	91.9	92.4
poly	75.4	84.2	89.2	95

Figure 24. Accuracy in assessment modules. MONO refers to items without translation ambiguity and POLY to items with translation ambiguity. The participants did the test on four occasions. A test was taken immediately after each of the four learning modules. In the immediate post-test (4), accuracy for items with two translation equivalents in the L1 is higher than for items with a single translation equivalent. This is likely associated with noticing.

In Study 1, CLI did not manifest itself in accuracy. Based on Figure 24 above, there seems to be a somewhat larger difference between the item types in Study 2. To test this, a generalized linear mixed effect model was performed to assess the development of accuracy in assessment across the two types of items. Random intercepts were included for PARTICIPANT and ITEM. TIME and ITEM TYPE were added as interactions. Once again, in terms of predictions, an effect for ITEM TYPE entails CLI and an interaction between TIME and ITEM TYPE that this effect changes over the time course of learning. WM, CC, I-CLSA, and E-CLSA were added as co-variates in line with the second and third research questions. The model is presented in Table 20 below. While statistically significant effects of forward CLI were found for processing cost in learning, CLI does not manifest itself in accuracy in Study 2 either.

With respect to aptitude in forward CLI, neither WM nor CC had any statistically significant effects in the models for processing cost or accuracy. However, there is a near significant effect of CC on accuracy in assessment with an estimate of -0.36470 ($p = 0.069$) on a scale from 0 to 1. For CC, negative values represent better CC whereas for accuracy, a higher score is better. An effect of CC on accuracy would entail that the participants are better at regulating influence in the L3 Kontu v2 from the L1 Swedish and the L2 English.

Table 20

Forward cross-linguistic influence in assessed accuracy over time

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-9.1180	0.1176	0.2202	0.3700	2.1857

Random effects

GROUPS	NAME	VARIANCE	SD
ITEM	INTERCEPT	0.8974	0.9473
PARTICIPANT	INTERCEPT	1.0706	1.0347

NUMBER OF OBS: 4284, PARTICIPANT, 30; ITEM, 72

Fixed effects

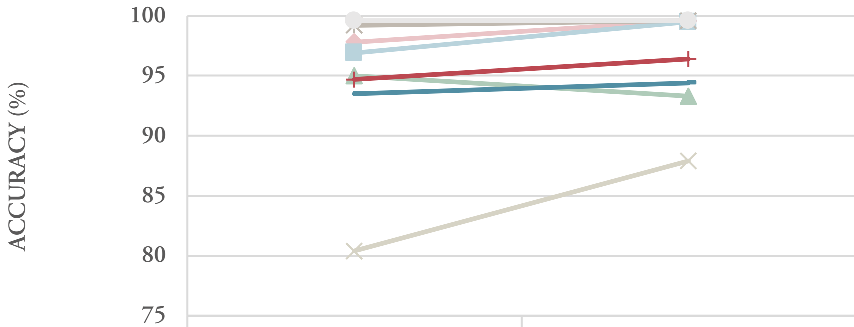
	ESTIMATE	SE	z value	*p*	
INTERCEPT	1.20482	0.34919	3.450	< 0.001	***
TIME (BLOCK)	0.60839	0.07103	8.565	< 0.001	***
ITEM TYPE	-0.15124	0.48996	-0.309	0.757	
WORKING MEMORY	0.16696	0.21592	0.773	0.439	
COGNITIVE CONTROL	-0.36470	0.20120	-1.813	0.069	.
E-CLSA PSYCHOTYPOL	0.30879	0.38200	0.808	0.419	
I-CLSA PSYCHOTYPOL	< -0.001	0.00208	-0.011	0.991	
TIME:ITEM TYPE	0.03829	0.14165	0.270	0.787	

Model fit
AIC 2754.3 BIC 2817.9 logLik -1367.1 deviance 2734.3 df.resid 4274

Note. Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod]. Performed in R 3.6.0 using using lme4 (1.1-23) glmerControl (optimizer = "bobyqa"). INTERCEPT is set to first assessment block, baseline items, mean working memory, mean cognitive control, and equal distance from Swedish and English for both E-CLSA and I-CLSA. WORKING MEMORY and COGNITIVE CONTROL have been normalized using Z-transformation. The only statistically significant effect is time (operationalized by progess in assessment blocks). Formula: glmer(accuracy ~ time*itemtype + scale(wm) + scale(cc) + (1|participant) + (1|event), family = "binomial", control = glmerControl(optimizer = "bobyqa")).

4.4.2 Reverse CLI in L1 Swedish and L2 English

The second possible direction of CLI in Study 2 is from the L3 Kontu v2 to the participants' L1 Swedish and L2 English. The primary interest is the development of the two types of pairs in the masked priming experiment: the baseline pairs (to control generic learning effects) and the experimental pairs that have translation ambiguity between the L1/L2 and the L3. The experimental pairs have been matched so that the translation ambiguity between the L3 and the L2, as well as between the L3 and the L1, is the same (to the extent it is possible to control given that the participants are native speakers of only one of the languages). Accuracy in pre- and post-test by item type is presented in Figure 25 below.

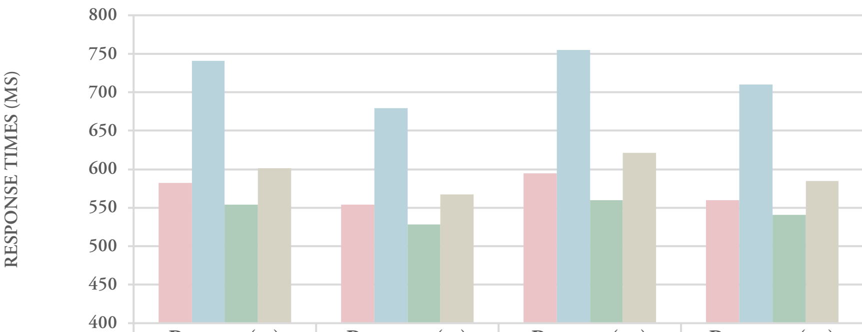


	pre	post
base (se)	97.8	99.6
base (en)	96.9	99.5
dist (se)	95	93.3
dist (en)	80.4	87.9
kontu (se)	99.2	99.6
kontu (en)	99.6	99.6
pseu (se)	94.7	96.4
pseu (en)	93.5	94.4

Figure 25. Accuracy in pre- and post-test conceptual priming tasks. 'Base' are translation equivalents of Kontu words (single translation equivalent in the L1/L2). 'Dist' are legal in L1/L2 phonotactics. 'Kontu' are related in Kontu but not in L1/L2. 'Pseu' are preceded by pseudowords in the mask.

The accuracy results in the L1 are similar to those in Study 1 except that the recognition accuracy for L1 Swedish distractors is higher. The English distractor items fared difficult for the L1 English participants in their first instantiation in Study 1 as well, and they seem to be even more difficult for the L2 English speakers. The item types that are used for the analysis of reverse CLI are “base” (both English and Swedish) and “kontu” (both English and Swedish).

The primary outcome variable was the development of a priming effect over time and by condition. The data is presented in Figure 26 below. What we can note is that there is a generic learning effect across all item types. Before adjusting for relative gains, both items with- and without translation ambiguity have faster RTs in the post-test.



	Pre-test (se)	Post-test (se)	Pre-test (en)	Post-test (en)
BASE	582	554	595	560
DIST	741	679	755	710
KONTU	554	528	560	541
PSEU	601	567	621	585

Figure 26. Visualization of RTs by condition and time. No adjustment for generic learning effects have been done. Data represents is mean response times in milliseconds. Table I7 in Appendix I presents more data on RTs in priming by item type and time. BASE are translation equivalents of Kontu words (single translation equivalent in the L1/L2). DIST are legal in L1/L2 phonotactics. KONTU are related in Kontu but not in L1/L2. PSEU are preceded by pseudowords in the mask.

In Study 1 we saw an initial increase in items with translation ambiguity in relation to the baseline items (i.e., after adjusting for generic learning effects). Figure 27 below shows that in the present study we see a similar pattern in the post-test for both the L1 Swedish and the L2 English. The L2 English is more affected than the L1 Swedish.

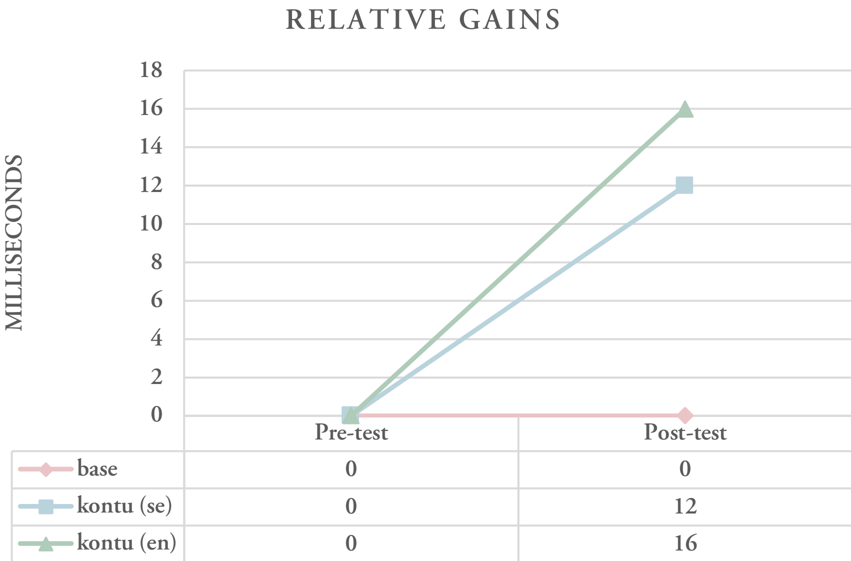


Figure 27. Visualization of data for assessing the relationship between learning effects and reverse CLI. Pre-test results are centered and baseline results kept constant throughout by deducting changes in those from all item types. We can see that Kontu items have higher response times (in comparison to baseline items) in the immediate post-test. This would align with the assumption of increased processing cost in the immediate post-test. Presented differences are in milliseconds.

A linear mixed effect model (presented in Table 21 below) was conducted to test reverse CLI from the L3 on the L1 and the L2. Random intercepts were added for PARTICIPANT and ITEM. TIME, ITEM TYPE, and LANGUAGE were added as interactions. WM, CC, E-CLSA, and I-CLSA were added as co-variates.

Table 21

Reverse cross-linguistic influence in priming over time

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.8574	-0.5876	-0.1898	0.3584	6.7009

Random effects

GROUPS	NAME	VARIANCE	SD
ITEM	INTERCEPT	832.4	28.85
PARTICIPANT	INTERCEPT	1949.3	44.15
RESIDUAL		15805.5	125.72

NUMBER OF OBS: 7502, PARTICIPANT, 30; ITEM, 64

Fixed effects

	ESTIMATE	SE	df	t	*p*	
INTERCEPT	650.44227	13.458	59.2819	48.331	< 0.001	***
TIME (BLOCK)	-39.6212	4.4520	7407.81	8.900	< 0.001	***
ITEM TYPE	-53.64816	22.777	821.442	-2.355	0.0187	*
LANGUAGE	-18.41613	9.92134	7407.09	-1.856	0.0635	.
WORKING MEMORY	-6.04313	8.91864	25.0296	-0.678	0.5042	
COGNITIVE CONTROL	27.4166	8.19466	25.0451	3.346	0.0026	**
E-CLSA PSYCHOTYPOL	1.68008	15.9664	24.9782	0.105	0.9170	
I-CLSA PSYCHOTYPOL	0.19238	0.08641	24.9961	2.226	0.0352	*
TIME:ITEM TYPE	17.95615	2.5158	7403.04	1.435	0.1514	
TIME:ITYP:LANGUAGE	-5.17192	17.5540	7403.08	-0.295	0.7682	

REML criterion at convergence 93973.5

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to pre-test, baseline items, mean working memory, mean retention, mean cognitive control, balansed perceived distance for both E-CLSA and I-CLSA. WORKING MEMORY, and COGNITIVE CONTROL have been normalized using Z-transformation. Formula: rt ~ time * itemtype + scale(wm) + scale(cc) + I-CLSA + E-CLSA + (1|item) + (1|participant).

There are statistically significant main effects (see Table 21 above) for TIME (Estimate = -39ms for post-test, p = < 0.001) and ITEM TYPE (Estimate = -53ms for items related in Kontu, p = < 0.01), as well as an effect nearing significance for LANGUAGE (Estimate = 18ms faster in Swedish, p = < 0.06). The interaction between TIME, ITEM TYPE, and LANGUAGE (which would have indicated reverse CLI and a difference between the quantity of reverse CLI in L1 and L2, respectively) was not statistically significant (Estimate = -5ms, p = < 0.77). The relative gains presented in Figure 27 above are visible in the model with an interaction for TIME and ITEM TYPE at 18ms as well as TIME, ITEM TYPE, and LANGUAGE at -5ms. CC has a significant effect (Estimate = 27ms, p = < 0.01) on response times. For CC, a lower score in the flanker task is better and better scores have lower estimates of response times in the priming task. This entails that participants with better CC were also overall faster in the priming task.

4.4.3 Psychotypology

Psychotypology was measured on two occasions: 1) using a conscious, explicit language similarity questions (E-CLSA) before learning Kontu v2, and 2) using unconscious, implicit assessment (I-CLSA), testing the participants' willingness to accept Swedish- and English-derived novel items in Kontu v2. The former task represents assumed similarity (before learning Kontu v2) and the latter perceived similarity.

In E-CLSA, Swedish and English were deemed to be most similar (M = 3.92 across all statements), followed by Swedish and Finnish (M = 2.63), which were deemed to be more similar than English and Finnish (M = 2.23) particularly with respect to vocabulary. Table 22 below presents the mean evaluations of similarity by linguistic sub-component. Some of the statements related to language similarity in whole while others related to a specific area of language similarity. A complete list of the statements is provided in Appendix H.

Table 22

E-CLSA: Mean evaluations of language similarity before learning Kontu v2 by component

	GLOBAL	GRAMMAR	IDIOM	PHON	SPEL	VOCAB
EN-FI	1.86	2.83	2.77	1.98	2.44	2.10
SE-FI	1.95	2.90	2.97	2.59	2.82	2.90
EN-SE	4.09	4.17	3.87	3.64	3.92	4.52

Note. Mean scores across participants. Values on seven-point Likert-scale. Higher value means more perceived similarity. Evaluated components: global similarity, grammar, translatable idioms and multi-word units, phonology and pronounciation, spelling, and vocabulary.

While in the E-CLSA conscious similarity evaluations L1 Swedish was deemed to be more suitable transfer source, in terms of performance in the I-CLSA which the participants took after learning Kontu v2, the participants seemed to be more willing to accept items from the L2 English. Novel English-derived items were accepted in 80.4% of the cases, whereas novel Swedish-derived items were accepted in 73.8% of the cases. Familiar Kontu v2 words were accepted in 98.3% of the cases and non-words violating Finnish phonotactics were rejected in 86.3% of the cases.

In addition to binary acceptability judgements, RT data was recorded and is presented in Table 23 below. Response times for L1 Swedish-derived items were higher than those for L2 English-derived items. This means that the participants were both less willing to accept L1 Swedish-derived items but also slower in making their assessment for L1 Swedish-derived items than for L2 English-derived items.

Table 23

I-CLSA: Unconscious psychotypology assessment (response times in ms by item type)

	N	*M*	*MDN*	*SD*	MIN	MAX
NOVEL SWEDISH	226	763	727	208	337	1366
NOVEL ENGLISH	232	708	670	229	336	1378
FAMILIAR KONTU	944	592	562	176	206	1392
NOVEL NON-WORD	235	760	731	186	220	1376

Note. Novel Swedish and Novel English refer to unfamiliar words in the target language that have have been derived from Swedish and English, respectively. Familiar items have been taught in the learning component. Non-words are novel items that violate Finnish (and Kontu v2) phonotactics.

Neither explicit nor implicit estimates of psychotypology had statistically significant impact on performance for forward CLI, although based on the estimates, the participants were more reliant of explicit (assumed) psychotypology. For reverse CLI (see Table 21 in Section 4.4.2), the effect of implicit (perceived) CLI was significant (p = 0.04).

4.5 Discussion

The main purpose of the present study was to replicate Study 1 with multilingual learners who had already acquired a second language and for whom Kontu v2 was a third language. It was hypothesized that learners who had already managed to acquire two languages would have better metalinguistic understanding of cross-linguistic similarity and ambiguity, be better at dealing with similarity and ambiguity cognitively, and would have managed to acquire language regulation skills in a way the *ab initio* monolingual participants in Study 1 would not have. Both accuracy and processing were tested for forward CLI towards the L3 (Kontu v2) whereas only processing effects were tested for reverse CLI as the assumption was that the learners would not make overt errors in their L1 or the L2 due to approximately one hour of instruction in the L3. Furthermore, for reverse CLI, the study attempted to estimate whether the L2 would be less resistant for cross-language activation of newly acquired vocabulary items than the L1.

The first research question related to the *direction of CLI*. Statistically significant effects of forward CLI were found but these were not in line with the findings of Study 1. The differences are outlined below. Furthermore, with respect to reverse CLI, in line with Study 1, no statistically significant effects of reverse CLI were found. Potential explanations for this are discussed below.

With respect to *forward CLI*, and in line with the results in Study 1, no statistically significant effects in accuracy were found for forward CLI in the present study. In Study 1, with *ab initio* monolingual learners, processing effects of translation ambiguity were observed in the learning outcome modules (LOTs), while in the present study they were observed primarily in learning. That is, the participants in Study 1 faced difficulty resolving translation ambiguity in tasks where they had to apply acquired knowledge. On the other hand, the participants in the present study faced difficulty resolving translation ambiguity in learning which resulted in better learning outcomes. There are two potential explanations for this.

Either, 1) the participants' better metalinguistic awareness (having acquired two languages) led to better learning outcomes in the LOTs, or 2) the results represent an effect of *desirable difficulty* in learning (Bjork, 1994). The latter refers to an effect that some increase in difficulty during learning leads to better end-result learning outcomes. In terms of the predictions of the PM, the results would suggest that the L3 learners were able to detect differences, rehearse new patterns already during learning, and revise the representations without consolidation taking place in between. From this perspective, the learners in the present study can be postulated to have been 'better' language learners than those in Study 1.

With respect to *reverse CLI*, no predictions are made by the PM. In the introduction, three explanations for negative reverse CLI (causing *slower* processing) and one explanation for positive reverse CLI (causing *facilitated* processing) were offered. It was postulated that two of the explanations of negative CLI as well as the explanation for positive CLI could at least theoretically take place at the early stages. The Inhibitory Control model (Green, 1986, 1998) would predict interference effects due to co-activation and necessary regulation of the learner's languages. Indications of such effects were found in the (immediate) post-test in Study 1 but they were not statistically significant. The same was found in the present study. The magnitude of (the statistically non-significant effects) was larger for the L2 than for the L1. The results could, equally well, be expected by the hypothesis by Levy, McVeigh, Marful and Anderson (2007) that retrieval of Ln words impairs their L1 equivalents. Given that the representations in the L3 are not particularly well established at this stage, but that on the other hand the amount of repetition of the items – particularly those of low frequency – was comparatively very high, the latter explanation would probably be more likely at this stage if any significant effect had been found. In Study 1, (statistically non-significant) facilitative effects were only found after a consolidation period. The present study did not include a consolidation period. Thus, it is impossible to say whether the facilitative effect would have appeared in the experienced language learners in the present study as well.

From a methodological perspective, there were multiple reasons for removing the consolidation and the delayed post-test from the present study, all of which were practical in nature. The addition of Swedish as an additional language meant that each priming task had to be administered in two languages, while the introduction of psychotypology meant two additional tasks, both adding to the length of the experiment. The strict restrictions on pre-existing L3s seriously limited the participant pool (students in Lund typically have either an additional mother tongue or have learned an additional language, like Spanish at school even if the proficiency is not always very high). Furthermore, less financial incentives in the form of compensation were offered to the participants in Study 2 (compared to Study 1). The most difficult part with recruitment in Study 1 was to convince participants to come to data collection on two occasions with a specific amount of time in between. Also, experiences from Study 1 suggest that very specific restrictions should be set in place with respect to what activities the participants are allowed to do between the two data collection sessions to ensure good quality sleep. Lastly, a good design of consolidation would entail two participant groups, each with the same amount of time between the two data collection sessions, but with one group having this time during the day and the other overnight. This would have doubled the required amount of participants. Given that the (non-significant) results in Study 1 suggest that reverse CLI seems to operate differently before and after consolidation, it would of course have been optimal to be able to explore this in a multilingual population as well. This remains, however, a topic for further research.

In line with the second research question, *aptitude* was also an aspect of interest in the present study. In Study 1, the participants were capable of acquiring Kontu v1 successfully irrespective of aptitude. There was, however, an effect of WM for the participants' ability to resolve translation ambiguity. No effects for CC were found. It was hypothesized that that this might be related to the participants being both overall novel language learners as well as initial state learners of Kontu v1, meaning that they would not have had to previously acquire skills to regulate cross-language activation. In the present study, neither WM nor CC predicted performance during learning. However, CC had a statistically significant effect in the reverse CLI task. This aligns with the assumption that the lack of modulation by CC was a result of the *ab initio* monolingual learners not having acquired language regulation skills, which the participants in the present study would have had to do before learning Kontu v2.

Finally, *psychotypology* was a new variable in the present study. Psychotypology was tested both though explicit and implicit means. E-CLSA was administered before learning Kontu v2 using Finnish as an approximant, while the I-CLSA was administered after learning Kontu v2. The scores in the two estimates of psychotypology have a low correlation coefficient ($r = 0.170$, $p = 0.361$). As mentioned before, the participants were left under the impression that they were acquiring a simplified variant of Finnish (which at the surface level is true). In the E-CLSA task before learning Kontu v2, the participants assessed Swedish to be more similar, which given the amount of shared lexical items across Swedish and Finnish is expected. However, while testing from which language the participants would be more willing to infer from, the participants were more prone to accept items from their L2 English. Neither approximation of psychotypology predicted accuracy or processing in learning. Perceived similarity (from the I-CLSA) modulated performance in the reverse CLI task with higher perceived relative similarity between English and Kontu v2 resulting in increased overall reverse CLI.

In hindsight, it would have been optimal to administer the E-CLSA on two occasions, before and after learning Kontu v2. This would have allowed for estimating whether the explicit evaluations of psychotypology had changed as a result of input in the TL. Of course, in the most optimal situation the I-CLSA would have been administered before any learning as well. However, evaluating *perceived* similarity before any input would be difficult, if not impossible. However, for societal reasons (with Finnish being widely spoken in Sweden) it can be assumed that the L1 Swedish learners, while pre-screened for not having any productive capacity in Finnish, have had a fair amount of input in Finnish from the surrounding society.

4.6 Conclusion

The main purpose of the study was to replicate Study 1 with learners for whom Kontu v2 is their third language. For this purpose, a new version of the pseudolanguage, Kontu v2, was created to adjust for any changes needed due to the addition of Swedish in the experiment. Furthermore, two tests of *psychotypology* were created: one explicit (E-CLSA) and one implicit (I-CLSA). These were administered before and after the learning of Kontu v2, respectively. The two tests estimated *assumed* and *perceived* similarity across the three languages.

Like in Study 1, no statistically significant effects of cross-linguistic influence (CLI) were found in outcomes. However, statistically significant effects were found in processing, highlighting the importance of including unconscious lexical activity when estimating CLI. The learners in the present study, compared to the learners in Study 1, faced more comparative difficulty (across item types) in learning. They, however, outperformed the participants in Study 1 in outcomes. It was hypothesized that this might either be due to better *metalinguistic skills* (Ringbom, 2007) or *desirable difficulty* (Bjork, 1994). Controlling for general learning effects, reverse CLI did present itself in the data but in quantities that fall well within the margin of error. In the magnitudes, the L2 was more effected than the L1, but the difference was so small that no major conclusions should be drawn from these findings.

Neither *working memory* (WM) nor *cognitive control* (CC) showed significant effects during learning, but better CC led to faster responses in the reverse CLI task. Furthermore, neither of the psychotypology measures had a significant modulating effect on learning. The I-CLSA psychotypology scores had a statistically significant effect on performance in the reverse CLI task.

5 Longitudinal Aspects of Naturalistic L3 Lexical Acquisition

5.1 Introduction

Studies 1 and 2 (Chapters 3 and 4 in this dissertation) looked at cross-linguistic influence (CLI) in the multilingual mental lexicon (MML) at the very initial stages of second- and third language acquisition (SLA and TLA, respectively). Meaning-based CLI was found in processing, i.e., time spent on task (ToTs) but not in accuracy, i.e., the proportion of correct and incorrect responses. Both forward and reverse CLI were of interest. In both studies, statistically significant effects were found for forward but not for reverse CLI. Furthermore, artificial language learning was used in both studies. The present study expands on the previous studies in three areas: 1) it presents longitudinal data, i.e., following a group of participants from a low level of proficiency to a high level of proficiency, 2) it presents data from naturalistic language learning, and 3) it presents data on all six potential directions of CLI in TLA.

In the present study, the primary point of interest lies in change over the course of learning. The assumption is that language learning entails fluctuating competence (Ecke, 2015). Longitudinal data was collected for this purpose from German (L1) and English (L2) speaking learners of Swedish (L3). The participants were followed from a beginner to a very high proficiency (CEFR C1 or higher) in Swedish. Change was measured, as in the previous studies, both in learners' perceptions of word meaning across the three languages as well as in time spent on task (ToTs) representing, among other things, difficulty of processing translation ambiguity. The development of all three aforementioned languages was investigated. Proficiency in the L3 in the present study was primarily operationalized as progress in the intensive language program.

Since all six potential directions of CLI are analyzed in the present study, there are three languages *in which* CLI can be observed and, in each of the three cases, there are two languages *from which* CLI can be observed. The language in which CLI is observed, i.e., L1 German, L2 English, or L3 Swedish, is referred to as the target language (TL) and the language *from which* CLI is observed, i.e., L1 German, L2 English, or L3 Swedish, is referred to as the source language (SL).[49] To give an example, in the case of forward CLI in L3 Swedish, the TL is the L3 Swedish, and the SLs are the L1 German and the L2 English. All statistical modeling has been performed so that there is always one TL and two potential SLs. To put it in other words, it is the participant's performance in the TL over the course of the longitudinal data collection that is being investigated in the present study. The SL-TL combinations are visualized in Table 25 in Section 5.3.2.

With respect to forward CLI, predictions can be made based on the *Parasitic Model* (PM) as well as the *Revised Hierarchical Model of Translation Ambiguity* (RHM-TA). The PM (Hall & Ecke, 2003) was also used to make predictions in Studies 1 and 2, and the model's predictions (presented in Table 3 in Section 2.2.2)[50] were found accurate for the tested initial stages. The model's predictions are less straightforward for the later stages of L3 lexical acquisition, but these are presented in Section 5.2. Furthermore, while the RHM-TA (Eddington & Tokowicz, 2013) only makes predictions for two languages, some of its predictions can be extended to L3 acquisition.

With respect to reverse CLI, Schmid & Köpke (2017) have proposed that this type of multilingual language acquisition would lead to attrition in the mother tongue (see Section 2.1). Furthermore, Linck, Kroll and Sunderman (2009) as well as Opitz (2013) have found that there seems to be an initial "hit" in the L1

[49] The terms *target language* and *source language* are often used so that they refer to the languages based on order of acquisition. This presumably aligns with that, in most cases, only forward *or* reverse CLI is investigated. As the present study investigates multidirectional CLI, testing all potential directions in one and the same population, this means that the source of CLI can be both a previously or a later acquired language. To streamline the use of terminology, order of acquisition is not considered when referring to the source *from which* and the target *in which* CLI is being observed. Instead, order of acquisition should be evident from the use of L1, L2, and L3.

[50] When comparing the presented predictions in the present chapter and those offered in Table 3, note that the Parasitic Model is a developmental model. That is, its predictions are from the perspective of the L3. The PM, for example, predicts that a word form in the L3 is connected to the closest L1 or L2 representation. This, in the case of the present study, would lead to CLI from the L1 or the L2 depending on whether a L1 or L2 representation was chosen.

that is related to extreme shifts in language use, but that this effect subsides after a while. It is possible that such effects relate to the interaction between acquiring new language competence and regulating skills. It was hypothesized that the *ab initio* monolingual participants in Study 1 had not yet acquired the skills required for regulating cross-language activation (CLA). The participants in Study 2, on the other hand, had acquired those skills for their L1 and L2, which they spoke fluently. One of the interesting aspects of the present study is that the participants progress up to such a high proficiency in their L3 during their participation that they would presumably have to acquire the capacity to regulate their L3 as well.

The participants were tested four times over ten months. Furthermore, the modulating factors aptitude and psychotypology were of interest in the study.

5.2 Aim and predictions

The main aim of the present study was to observe the development of conceptual-lexical relationships in naturalistic language acquisition from a longitudinal perspective in learners who – during the participation – become highly proficient speakers of their L3. Given this aim, the following research questions have been formulated:

RQ1 If present, is cross-linguistic influence (CLI) in the multilingual mental lexicon (MML) unidirectional or multidirectional?

RQ2 To what extent is CLI in the MML affected by aptitude and psychotypology?

RQ3 Are the aforementioned aspects of directionality and the effect of modifying factors dependent on the stage of acquisition?

The first research question relates to the direction of CLI. Since there are three languages that are of interest, there are six potential directions of CLI. All of these are of interest, i.e., from the previously acquired languages (German L1, English L2) to the L3 (Swedish) and vice versa, as well as between the previously acquired languages. The second research question relates to factors that represent individual variation in language learners. Aptitude is operationalized as cognitive control (CC) in the present study. Psychotypology is treated as a conscious construct and is measured overtly. The third research question relates to the effect of stage of acquisition operationalized by progress (over four data points and over a period of ten months) in the intensive language course.

With respect to the *direction* of CLI, the point of departure is the assumption that a speaker's languages are not functionally separate (Kroll, Dussias, Bice & Perrotti, 2015). Given that the participants become highly proficient speakers of the L3 during their participation, the expectation is that we see some form of CLI in all directions, presumably modulated by proficiency, possibly by available cognitive resources (modulated by aptitude), and psychotypology. The study takes a longitudinal approach, meaning that the initial performance of the participants acts as their baseline. The present study also represents the first set of data from a broad range of proficiency levels in this dissertation.

The *Parasitic Model* (PM) by Hall and Ecke (2003) was used to make predictions in Studies 1 and 2 and the model's predictions (presented in Table 3 in Section 2.2.2) were found accurate for the tested initial stages. In the case of the present study, the PM only makes predictions for three of the potential six directions of CLI. These are the influence of the L1 on the L3, the L2 on the L3, and (with slight adaptations, like in Study 1) the L1 on the L2. Since the participants in the present study have already acquired the L2 English to a very high level of proficiency, no changes would be expected at the language level (as opposed to the item level) in terms of L1 to L2 influence with the exception of further automatization of access routes between the lemma and the lexeme (see Section 2.2). This means that it is the predictions of the former two directions, i.e., L1 to L3 and L2 to L3, that are of interest in the present study. The PM predicts that a word form in the L3 is connected to either a L1 or a L2 representation depending on which is deemed to be more similar. Unlike any of the other developmental models presented in Chapter 2 in this dissertation, the PM opens up for the possibility that in the case no matching translation equivalent in the L1 *or* the L2 can be identified, a separate L3 frame that connects directly to a conceptual representation will be formed. Furthermore, the PM suggests that in a process of revisiting, bypassing, severing, and refining the representation of individual words, the connections become more autonomous. This aligns with the L1 Lemma Mediation Hypothesis (Jiang, 2002) and the Revised Hierarchical Model (RHM) by Kroll and Stewart (1994) as well as the RHMs more recent modifications. However, the PM predicts fluctuation rather than a linear fashion of increasing independence.

In terms of *method*, to facilitate analysis of meaning-based CLI, the instruments in this study (presented in Section 5.3.2 below), as opposed to those in Studies 1 and 2, no longer test learning outcomes operationalized by accuracy and processing cost of individual words. Instead, the outcome variable is now perceived similarity between two words (and associated processing cost) and the change of the evaluation over the course of learning. This task was not strictly timed, meaning that the learners can utilize *introspection* (see Section 2.3.1), which refers to Jiang's (2002) proposal that multilingual speakers deal with translation ambiguity using conscious rules. With the possibility of introspection, CLI will likely present itself in *either* evaluations or the time it takes to make those evaluations or as an interaction between the two. That is, the learners can, *if* they are aware of the dissimilarities between particular words (see the predictions of the PM) between the source language and the target language, adjust their similarity evaluations in the target language, resulting in similarity evaluations that are not affected (or are less affected) by CLI. However, in these cases we should be able to observe the presence of CLI in the time it takes to evaluate the similarity of the two words instead.

The first modifying factor in the present study was *aptitude* (operationalized by cognitive control). In Studies 1 and 2, which included a learning component, working memory (WM) was also a part of the operationalization of aptitude as it was hypothesized to predict performance in learning. In the present study, language learning, *per se*, is not an outcome variable. Thus, WM was not measured. However, with respect to CC, the assumption is that it will play a more important role than in Studies 1 and 2. Since the participants in the present study are multilinguals with (at the end of their participation) a very high proficiency in three languages, the hypothesis is that CC should aid the learners in regulating the L1, the L2, and the L3, potentially leading to less CLA and, subsequently, less CLI.

The second modifying factor in the present study was *psychotypology*. The assumption is that if the participants use *introspection* to resolve translation ambiguity, we should see a longer time on task (ToTs) for the items with translation ambiguity. On the other hand, if the participants rely on *global inhibition* of the previously acquired languages as a learning strategy, as suggested by Higby et al. (2020), then we could hypothesize that the quantity of the inhibition is modulated by the amount of perceived language similarity.

5.3 Method

5.3.1 Participants

The participants were native speakers[51] of German who spoke German at home in childhood, at school, and grew up at a predominantly German-speaking environment. They had substantial previous knowledge of English (CEFR B2 or higher, but not native). The participants were taking part in intensive language training in Swedish at Lund University in Sweden. All participants fulfilled the conditions of admission to Swedish universities, including the English-language proficiency requirements. The participants belonged to different cohorts, as data was collected from students in an intensive language program between 2016 and 2019.

The participants filled in four background questionnaires during the data collection period: one for each data collection point. The first questionnaire was more comprehensive, while the other three focused on assessing matters that could have changed throughout the learning period, such as changes in relative language use, attitude, and self-evaluated proficiency. The questionnaires, both of which are available in the Appendix (J and K), were developed partly based on the LHQ 2.0 questionnaire by Li, Zhang, Tsai, and Puls (2014) as well as the one used in Suhonen (2015). The questionnaires were administered in a pen-and-paper format and before any other data collection in all data collection sessions.

There were a total of 8 participants, ranging from 20 to 30 in age ($M = 24.8$, $SD = 3.73$). Of the participants, seven (88%) were female and one (12%) male. The participants all had high school-level education completed. The participants reported on average 24.8 years of use in the L1 German (Min-Max = 20-30), 13.9 years of use in the L2 English (Min-Max = 10-17), and on average 1.5 years of use in the L3 Swedish (Min-Max = 0.2-3) at the onset of the intensive program. Five of the participants completed the whole 10-month intensive language program. Two of the participants entered the program after it had already started since they already had some pre-existing knowledge of Swedish, and one participant left the program after five months.

[51] Monolingual, or based on different definitions of bilingualism, they could be equally well be considered to be bilingual in German and English (and perhaps other languages) despite a late age of onset past the mother tongue.

At the onset of data collection, the participants self-evaluated their learning aptitude (*M* = 5.25, Min-Max = 4-6, *SD* = 0.71), attitude (*M* = 6.25, Min-Max = 5-7, *SD* = 0.89), and ability for cross-cultural communication (*M* = 5.63, Min-Max = 4-7, *SD* = 0.92) on a seven-point Likert scale.

5.3.2 Instruments

The main instruments for data collection in Study 3 were the Eriksen Flanker task used to measure CC, the Explicit Cross-Linguistic Similarity Assessment (E-CLSA) that was used to measure psychotypology, a lexical placement module from DIALANG used as a proxy for proficiency, and a Word Pair Similarity Perception Task (WSPT) used to measure CLI. The Eriksen Flanker task was also used in Studies 1 and 2 (see 3.3.2.3). In the present study, aptitude is operationalized as CC and unlike in Studies 1 and 2, the Eriksen Flanker task (Eriksen and Eriksen, 1974) was presented in PsyToolkit (Stoet, 2010; 2017). E-CLSA was also used in Study 2 (see 4.3.2.2). The latter two tasks were introduced for Study 3. All four experimental instruments are visualized in Figure 28 below and the two new experimental instruments are presented below in this the present section.

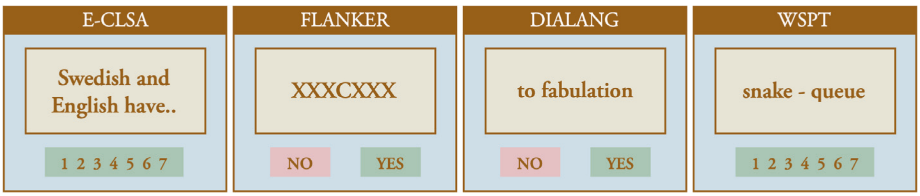


Figure 28. Experimental instruments used in Study 3: 1) Explicit Cross-Linguistic Similarity Assessment (E-CLSA) that was used to measure psychotypology, 2) Eriksen Flanker task used to measure CC, 3) lexical placement module from DIALANG used as a proxy for proficiency, and 4) Word Pair Similarity Perception Task used to measure CLI.

Multiple measures of *proficiency* were collected.[52] For each data collection point in the longitudinal study, the participants self-evaluated their language proficiency – faceted for listening, speaking, reading, and writing – on a seven-point Likert scale for *all* of their languages, including German, English, and Swedish, as well as any additional languages. The scale is illustrated in Table 24 below.

[52] Correlations between these are presented in Appendix N.

Table 24

Self-assessment scale of language competence

Very Poor	Poor	Limited	Functional	Good	Very Good	Native-like
1	2	3	4	5	6	7

Note. The seven-point Likert scale used for self-evaluations of language proficiency.

For the three languages of interest, the participants also evaluated their language skills in the four aforementioned areas using CEFR can-do type statements.[53] In addition to the self-assessments of language proficiency, a range of independent language proficiency measures were collected throughout the course of data collection. These represent classroom progression as well as vocabulary size measures and semi-elicited oral language production. Out of these, progress in the intensive program and vocabulary size were used as predictors in the present study. Correlations between the proficiency measures are presented in Appendix N.

One of the strengths of the chosen design is the possibility to map the participants' *progression* according to the Common European Framework (CEFR). The eight-module course 'Swedish as a Foreign Language 1-8,' 60 ECTS,[54] encompasses CEFR levels A1 to C1 with the final examination for module 8 being at the C1-level. The final examination for module 4 is at B1 level. The modules are taken consecutively. Furthermore, the students are required to pass an examination at the end of each module. The four data collection points in the present study align with modules 2, 4, 6, and 8. The eight and the last module has an exam that is certified at CEFR C1 and can be used to fulfill the Swedish requirements for entrance to university-level studies in Swedish.

The *progression* of the participants in the intensive Swedish course gives estimates on the increase of their Swedish-language proficiency. However, since vocabulary is a particular area of interest, an additional proficiency component – a vocabulary-based placement module from the DIALANG proficiency test[55] – was

[53] The CEFR can-do –statements take a user-oriented approach to the speaker's abilities, whereby the language user's abilities are defined as the ability to perform certain actions language use actions rather than perform at a certain level of correctness.

[54] ECTS refers to the European Credit Transfer and Accumulation System. Each credit requires 25-30 hours of work. 60 ECTS points correspond to one academic year (European Union, 2015).

[55] The DIALANG test was created by Meara et al. (Alderson, 2005). The lexical placement module has been found to be a reliable predictor of performance in other modules in the DIALANG test.

administered. In this task, the participants had to indicate whether 75 words were real words in the target language or not (50:25 correct to incorrect ratio). Points were awarded for correctly recognizing real words. Missing real words caused no consequences. Indicating pseudowords as real words resulted in points being deducted.

Data for the main outcome variable was collected using a similarity evaluation task. The participants evaluated similarity between two words within a language in a Word Pair Similarity Perception Task, henceforth abbreviated as WSPT. The same data was collected at all four data collection sessions. Evaluations from the first session function as a baseline in terms of progress. The WSPT comprises of 360 word pairs, which were presented to the participants on a computer screen using SuperLab 5 and the evaluations were captured using a Cedrus RB-740 response pad which has seven buttons. For the purposes of the experiment, the buttons were labeled with numbers from 1 to 7, forming a Likert scale. The WSPT was modeled after Jiang (2002), and while Jiang's version was designed to be used between two languages, the task was adjusted to be used for three languages based on an earlier experiment presented in Suhonen (2015).

There are 180 experimental items and 180 filler items in the WSPT. The experimental items are divided across six experimental conditions (C1 through C6, presented in Table 25 below) of 30 items each. The six conditions encompass all the potential combinations of the tree languages as a source, target, and a "nuisance" language.[56]

Table 25

Word Pair Similarity Perception Task Conditions

TARGET	GERMAN	ENGLISH	SWEDISH
SOURCE	[C1] English	[C3] German	[C5] German
	[C2] Swedish	[C4] Swedish	[C6] English

Note. The combinations C1 through C6 refer to the conditions in the data set and the appendix.

There are three types of items within the item set in each condition: *ambiguous* items with translation ambiguity, *baseline* items without translation ambiguity, and *fillers* (see examples and a walkthrough of items in Figure 29 below).

[56] *Nuisance language* in this context is the third language, i.e., the language that is neither the source or the receiver language in the theoretically assumed CLI.

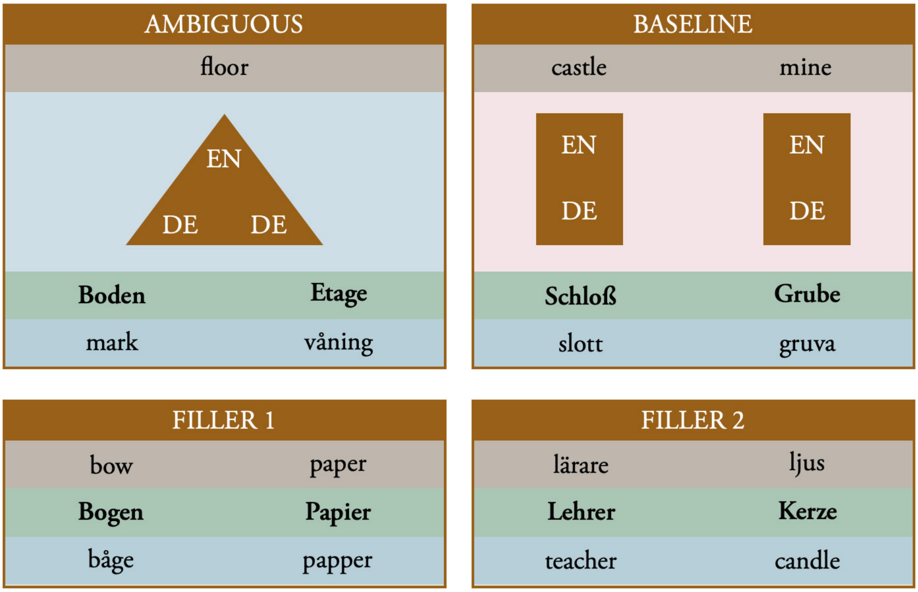


Figure 29. Word Pair Similarity Perception (WSPT) task item example from English to German [condition C1]. English on grey, German on green, and Swedish on blue background. Presented word pairs in bold. The two main item types are the words with translation ambiguity across languages and the baseline items. In the ambiguous-type items, two of the languages operate similarly while one of the languages operates differently. For the baseline items, all three languages operate similarly. In addition to these two types, two sets of fillers were created for each item type pair.

CLI is expected for the ambiguous items (n = 15), but not for the baseline items (n = 15). The amount of fillers is the same as the joint total amount of ambiguous and baseline items in each condition. This means that the stimuli for each condition consists of a total of 60 (15+15+30) word pairs. The items with translation ambiguity are noun pairs in the TL of the given condition with a single translation equivalent in the given SL for the two words in the TL. The baseline items are noun pairs just like the ambiguous items and have been matched for relationship between the two words in the TL and the SL as well as the individual words' frequency, and length. The filler items were also roughly, albeit not as stringently, matched for similarity in meaning, frequency, and length with the experimental items to ensure that the experimental items do not stand out from the set of items. The item creation process is outlined in Appendix L and the items are listed by condition in Appendix M.

In item creation, formal similarity was avoided to the extent it was possible (in three highly related languages) between the SL and the TL in the items with translation ambiguity and baseline items but not in fillers. In the example above, the English 'floor' refers to both German *Boden* and *Etage.* The baseline items, Schloß and Grube, are items where no CLI is expected since they translate to two items in both of the other languages. Additionally there are two filler items, in the present example Bogen – Papier, and Lehrer – Kerze.

5.3.3 Procedure

In the beginning of the first data collection session (DP1), the participants were informed about the purpose of the study and were allowed to ask questions. They were asked to sign a consent form. After it was mutually agreed that data collection could commence, the participants were asked to fill in a background questionnaire. The participants took all the tasks in one sitting per each data collection session.

The instruments in each data collection session were divided into two distinct categories: 1) the non-language specific instruments including the background questionnaire, the self-evaluations of language proficiency, the cognitive control, and the psychotypology instruments were all done in English, and 2) the language specific instruments including the six experimental conditions as well as the speech production tasks and the DIALANG vocabulary placement task for each language were done as single language units in the target language.

There were three blocks in the experimental instrument section, each consisting of only one language on the surface level, i.e., the participants were aware of only one language being used in each block. This means that within that block all instructions, as well as all presented material, were in the target language alone. However, within each block, two experimental conditions were presented where one of the two other (non-target) languages was the primary source of expected CLI. In the beginning of each block, the participants were presented with approximately five minutes of dialogue from The Moomins in the target language to tune the participants to the target language mode (Grosjean, 1982). The setup of the data collection points, as well as the background block and the three experimental blocks included in each data collection point are presented in Figure 30 below.

The order of languages in the blocks was randomized except that Swedish was always the last block in the first data collection session. All instructions were given in the target language in each block, and since at the time of the first data collection point, the participants' language skills were not necessarily at the level where they would be able to comprehend all the instructions or to be able to ask questions, placing the Swedish condition last allowed them to perform the tasks based on the comprehension of the instructions already from the previous blocks.

All in all, the data collection took approximately 180 minutes for the first session, 90 minutes for the subsequent two sessions, and 120 minutes for the last session. This represents the different lengths of the questionnaires administered at each data collection session as well as some generic familiarity effects with the instruments.

BACKGROUND ENGLISH	Welcome and consent[57] Background questionnaire Cognitive control Psychotypology
BLOCK 1 RANDOM LANGUAGE	A video in target language #1 Similarity evaluations: 15 treatment and 15 baseline items from two conditions with the same target language and 60 fillers – a total of 120 items. Break after 60 word pairs. Speech task DIALANG
BLOCK 2 RANDOM LANGUAGE	A video in target language #2 Similarity evaluations: 15 treatment and 15 baseline items from two conditions with the same target language and 60 fillers – a total of 120 items. Break after 60 word pairs. Speech task DIALANG
BLOCK 3 RANDOM LANGUAGE	A video in target language #3 Similarity evaluations: 15 treatment and 15 baseline items from two conditions with the same target language and 60 fillers – a total of 120 items. Break after 60 word pairs. Speech task DIALANG Post-experiment interview / debriefing

Figure 30. Procedure in Study 3. Each data point followed the same pattern even if not all types of data were collected on all occasions. Speech samples were collected for another study.

[57] None of the conditions by the Swedish Research Council that mandate ethical review apply for the present study. A privacy policy and data management description was provided to the participants. In accordance with GDPR, the European Union General Data Protection Regulation (2016/679), the participants consented in writing to the following: 1) that their personal information will be stored separately from research data, 2) that personal information will be stored in relation to their compensation in accordance with the Swedish accounting legislation (BFL 1999:1078) by the university centrally, and that 3) the collected but anonymized research data will be publicly available and may be used for further research purposes without additional consent from the participant, 4) and that it is possible to retract one's participation up until the point of publication. The participants were compensated hourly in by the university.

Table 26 lists which research instruments were administered at each data collection point. The participants were compensated hourly for their participation by the university. Furthermore, at most data collection sessions, a trilingual speaker of German, English, and Swedish was present in the premises for the purposes of acting as a dialogue partner for the speech production samples.

Table 26

Included tasks and instruments at different data collection points

INCLUDED TASKS	1	2	3	4
Long Questionnaire	X			
Short Questionnaire		X	X	X
Self-evaluations of Proficiency	X	X	X	X
E-CLSA Psychotypology	X			X
DIALANG Proficiency German	X			X
DIALANG Proficiency English	X	X	X	X
DIALANG Proficiency Swedish	X	X	X	X
Eriksen Flanker Cognitive Control	X			X
CLI WSPT Condition 1	X	X	X	X
CLI WSPT Condition 2	X	X	X	X
CLI WSPT Condition 3	X	X	X	X
CLI WSPT Condition 4	X	X	X	X
CLI WSPT Condition 5	X	X	X	X
CLI WSPT Condition 6	X	X	X	X

Note. The leftmost column presents the administered tasks and instruments. The four columns on the right present the four data collection points with included instruments marked on each row.

5.3.4 Data analysis, transformations, and structure

There were two outcome variables in Study 3: 1) WSPT similarity evaluations, representing meaning-based CLI in items with translation ambiguity, and 2) WSPT time spent on task (ToTs) representing the time it took to complete each similarity evaluation. All statistical analysis was performed in R (R Core Team, 2018) and data exploration was done using Jamovi (The Jamovi Project, 2019).

No transformations nor trimming were performed to WSPT or CC data. The data set for the WSPT similarity evaluations and WSPT ToT data contains 8,564 observations. The participants were asked to work as quickly as possible in the WSPT task, but the task did not time out and was designed to allow introspection, and thus it would not be optimal to do upper-spectrum trimming. The CC data consists of 800 observations and was not deemed to need trimming.

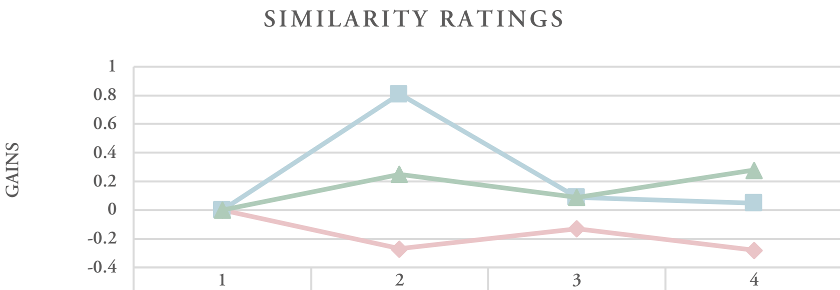
Psychotypology data was collected on two occasions: in the beginning and at the end of the longitudinal study. Global measures of psychotypology from the E-CLSA were computed from the area specific questions by giving equal weight for each question. The psychotypology scores for data points two and three were interpolated in a linear fashion from the first and the last data points.

Since the participants were native speakers of German, the German proficiency tests were administered only in the first and the last data collection session, while the English and Swedish proficiency tests were administered at all four data collection sessions. German language proficiency for data points two and three were interpolated in a linear fashion from the first and the last data points. English and Swedish, on the other hand, have proficiency data from all four data collection sessions.

5.4 Results

5.4.1 Forward CLI

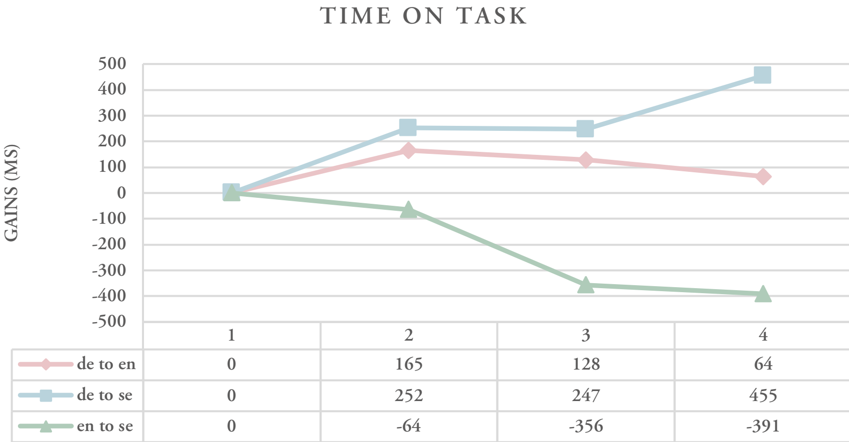
Forward CLI was measured using the WSPT task. Two measures of CLI were collected in the WSPT: 1) *similarity ratings* in items with translation ambiguity, and 2) *time spent on task* (TOTs). Change in both was measured in relation to change in the baseline items over time to compensate for generic learning and processing effects. Zero represents unchanged quantity of CLI relative to the quantity at first data point. Figure 31 below shows gains in similarity ratings in the three possible directions of forward CLI in the sample population. It should be noted that the language skills of the participants are not developing uniformly across the three languages. At time point one, the participants already have a high proficiency in L2 English, which presumably keeps developing, while L3 Swedish skills are still relatively low at around CEFR A2. By the fourth data collection session, the participants are at least at CEFR C1 in both non-native languages.



	1	2	3	4
de to en	0	-0.27	-0.13	-0.28
de to se	0	0.81	0.09	0.05
en to se	0	0.25	0.09	0.28

Figure 31. WSPT similarity ratings in forward CLI (1-7 Likert) over the four data points (DP). The scores represent gains over time. All scores have been centered to align with baseline items at DP1 and subsequent scores represent gains from DP1 corrected for changes in the baseline items.

What we can note from the data in Figure 31 above is, that the amount of CLI measured from the similarity ratings is relatively stable from L2 English to L3 Swedish throughout learning. There is a slight increasing trend and some fluctuation. Influence from L1 German to L3 Swedish initially increases (by 0.8 points on a 7-point scale) and then decreases to roughly the same as than in the beginning of the data collection period. The effect of L1 German to L2 English has a downward trend.



	1	2	3	4
de to en	0	165	128	64
de to se	0	252	247	455
en to se	0	-64	-356	-391

Figure 32. WSPT time on task in forward CLI (ms) over the four data points (DP). The scores represent gains over time. All scores have been centered to align with baseline items at DP1 and subsequent scores represent gains from DP1 corrected for changes in the baseline items.

In terms of *processing*, we can observe (a pattern) in Figure 32 that there is an ever-increasing attempt (or cost) to inhibit forward CLI from L1 German to L3 Swedish at the item level for items with translation ambiguity. The same effect can initially be seen in L2 English as well. On the other hand, there is a reducing trend for processing cost for items with translation ambiguity between L2 English and L3 Swedish. Remember that the participants already had a relatively high proficiency in their L2 at the commencement of the data collection.

Comparing the WSPT similarity ratings and ToTs for forward CLI from the L1 German to the L2 English, we can note that there is first a reduction in the amount of CLI in the similarity ratings which coincides with increased ToTs. For CLI from L1 German to L3 Swedish, there is an initial increase which then subsides but the ToTs keep increasing. For CLI from the L2 English to the L3 Swedish, there is a downward trend in ToTs throughout the data collection period. The similarity evaluations fluctuate but there is presumably a trend towards increasing CLI. Considering the ToTs and similarity evaluations, it seems that there is an increased preference over relying on (or accepting) English while using Swedish as opposed to relying on (or accepting) German at the item level. Whether this is a result of conscious or unconscious processing, or a result of changes in cross-linguistic activation cannot be deducted with certainty from the data. However, the trend for the quantity of influence observed in the similarity ratings is increasing for English and reducing for German. The pattern for inhibition as observed in ToTs is the opposite.

Two linear mixed effect models (LMEs) were performed to test the statistical validity of the aforementioned observations: one with WSPT similarity ratings as the outcome variable and one with WSPT ToTs as the outcome variable. Random intercepts were included for PARTICIPANT and ITEM. The main (fixed) factors TIME, ITEM TYPE, and SL were added as interactions. Additionally, TOT was added as an interaction in the model with WSPT ratings as the outcome variable. Modifying factors CC, E-CLSA, SL PROFICIENCY, and TL PROFICIENCY were added as co-variates in line with the research questions. The TL is the L3 Swedish and the SLs are the L1 German and the L2 English. The models are presented in Table 27 and Table 28 below.

Table 27

Forward CLI in Swedish from German and English (WSPT similarity ratings)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.8032	-0.6370	0.0212	0.6253	2.9337

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	1.7118	1.3084
PARTICIPANT	INTERCEPT	0.3867	0.6218
RESIDUAL		1.7461	1.3214

NUMBER OF OBS: 1291, PARTICIPANT, 8; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	3.579e+00	8.854e-01	4.949e+02	4.042	< 0.001	***
TIME	-1.577e-01	8.449e-02	1.176e+03	-1.866	0.0622	.
ITEM TYPE	-7.904e-01	5.566e-01	9.345e+01	-1.420	0.1589	
SOURCE LANGUAGE	-4.459e-01	5.863e-01	1.145e+02	-0.761	0.4485	
TIME ON TASK	-9.372e-02	2.685e-01	1.215e+03	-0.349	0.7271	
COGNITIVE CONTROL	1.949e-01	2.672e-01	6.396e+00	0.730	0.4915	
E-CLSA PSYCHOTYPOL	2.271e-02	6.744e-02	1.212e+03	0.337	0.7363	
SL PROFICIENCY	8.449e-05	6.693e-04	1.090e+03	0.126	0.8996	
TL PROFICIENCY	1.227e-03	6.603e-04	7.761e+02	1.858	0.0636	.
ITEM TYPE:TOT	8.838e-01	3.427e-01	1.214e+03	2.579	0.0100	*
TIME:ITYP:TOT	-2.164e-01	1.128e-01	1.212e+03	-1.918	0.0554	.
TIME:ITYP:SLANG:TOT	9.120e-02	1.571e-01	1.213e+03	0.580	0.5618	

REML criterion at convergence 4644.7

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). SL = L1 German or L2 English. TL = L3 Swedish. INTERCEPT is set to time 1, baseline items, source language German, as well as mean CC, ToTs, psychotypology, SL proficiency, and TL proficiency. RESPONSE TIMES and COGNITIVE CONTROL have been normalized using Z-transformation. Formula: rating ~ time * itemtype * source language * scale(ToT) + scale(cc) + E-CLSA + SLP DIALANG + TLP DIALANG + (1|item) + (1|participant).

Table 28

Forward CLI in Swedish from German and English (WSPT ToTs)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.3217	-0.5194	-0.0986	0.3315	9.4047

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	86941	294.9
PARTICIPANT	INTERCEPT	411854	641.8
RESIDUAL		1072182	1035.5

NUMBER OF OBS: 1291, PARTICIPANT, 8; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	2260.3412	632.4474	247.4413	3.574	< 0.001	***
TIME	-48.1643	65.6265	1214.1295	-0.734	0.4631	
ITEM TYPE	-55.9902	246.1394	493.7307	-0.227	0.8202	
SOURCE LANGUAGE	-188.6955	281.1278	695.0803	-0.671	0.5023	
COGNITIVE CONTROL	462.9069	270.5526	6.1812	1.711	0.1365	
E-CLSA PSYCHOTYPOL	60.5031	49.0987	1222.7521	1.232	0.2181	
SL PROFICIENCY	0.6197	0.5106	1187.6343	1.214	0.2251	
TL PROFICIENCY	-0.8401	0.5191	1030.5843	-1.618	0.1057	
TIME:ITEM TYPE	136.0708	74.3462	1216.7738	1.830	0.0675	.
TIME:SLANGUAGE	168.3211	75.5143	1217.9892	2.229	0.0260	*
TIME:ITYP:SLANGUAGE	-280.1201	105.1947	1216.7751	-2.663	0.0079	**

REML criterion at convergence 21559.9

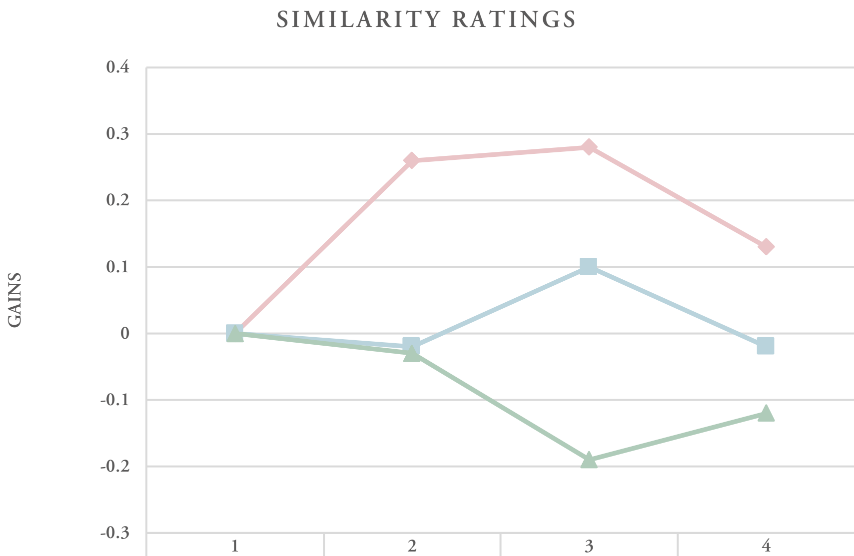
Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). SL = L1 German or L2 English. TL = L3 Swedish. INTERCEPT is set to time 1, baseline items, source language German, as well as mean CC, psychotypology, SL proficiency, and TL proficiency. COGNITIVE CONTROL has been normalized using Z-transformation. Formula: ToT ~ time * itemtype * source language + scale(cc) + E-CLSA + SLP DIALANG + TLP DIALANG + (1|item) + (1|participant).

In the results of the statistical models presented in Tables 27 and 28 above, forward CLI presents itself differently in the two outcome variables. For WSPT similarity ratings, we see a near-significant (both $p = 0.06$) generic learning effect and effect of proficiency in the L3 Swedish. Furthermore, we have a statistically significant interaction for item type and ToTs ($p = 0.01$) with higher response times for items with translation ambiguity resulting in lower similarity ratings, and near-significant interaction for time, item type, and response time ($p = 0.05$). There is no statistically significant difference of whether the translation ambiguity hails from the L1 German or the L2 English. For WSPT ToTs, we have a significant interaction ($p = < 0.01$) for the main effects time (data points 1 through 4), item type, and source language with L1 German resulting in more forward CLI at the item level and an increasing effect for L2 English at the item level over time.

5.4.2 Reverse CLI

Reverse CLI was measured using the WSPT task. Figure 33 below shows gains in similarity ratings and Figure 34 below gains in ToTs in the three possible directions of reverse CLI in the sample population. As noted earlier, the language skills of the participants are not developing uniformly across the three languages. Like with the analysis of forward CLI, the results represent change in comparison to the commencement of data collection, i.e., zero in value in the figures below does not represent a lack of CLI but rather no relative change in the quantity of CLI. At time point one, the participants already have a high proficiency in the L2 English, while L3 Swedish skills are still relatively low. The learning rate in the L3 Swedish is unusually fast, the quantity of daily use of the L3 Swedish is very high compared to typical classroom learners, and the learners use very little of their L1 German being immersed in the L3 Swedish environment in Sweden.

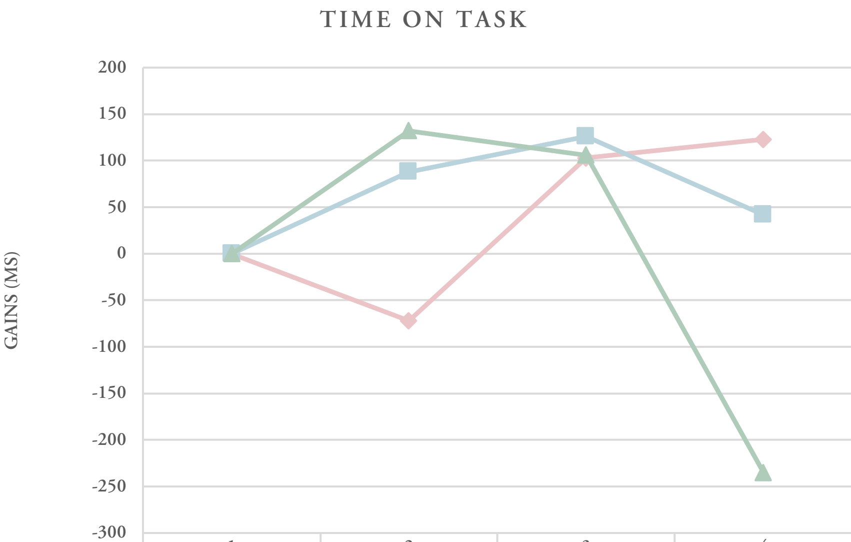
What we can note from Figure 33 below, is that the effect of the (already acquired) L2 English on the L1 German stays relatively stable in similarity ratings. Both L3 Swedish to L1 German CLI, and L2 English to L1 German CLI have a curvilinear relationship in similarity ratings that reverts back towards pre-Swedish-learning levels. In the case of L3 Swedish to L1 German reverse CLI, the effect over the course of learning is that of increase whereas with L3 Swedish to L2 English reverse CLI the pattern is the opposite.



	1	2	3	4
se to de	0	0.26	0.28	0.13
en to de	0	-0.02	0.1	-0.02
se to en	0	-0.03	-0.19	-0.12

Figure 33. WSPT similarity ratings in reverse CLI (1-7 Likert) over the four data points (DP). The scores represent gains over time. All scores have been centered to align with baseline items at DP1 and subsequent scores represent gains from DP1 corrected for changes in the baseline items.

Perhaps rather expectedly (given results from Studies 1 and 2, as well as forward CLI in the present study), we see an opposite pattern (than for similarity ratings) in *processing cost* in Figure 34 below. Reverse CLI from L3 Swedish to L2 English, with higher similarity ratings, is connected to longer ToTs in a similar curvilinear pattern, and reverse CLI from L3 Swedish to L2 English has the opposite pattern. Again, for ToTs, the overall magnitude of change in reverse CLI is smaller than for those observed for forward CLI.



	1	2	3	4
se to de	0	-72	103	123
en to de	0	88	126	42
se to en	0	132	106	-235

Figure 34. WSPT time on task in reverse CLI (ms) over the four data points (DP). The scores represent gains over time. All scores have been centered to align with baseline items at DP1 and subsequent scores represent gains from DP1 corrected for changes in the baseline items.

Two LMEs were performed to test the statistical validity of the aforementioned observations: the first with WSPT similarity ratings as the outcome variable and the second with WSPT ToTs as the outcome variable. Random intercepts were included for PARTICIPANT and ITEM. The main (fixed) factors TIME, ITEM TYPE, and SL were added as interactions. Additionally, TOT was added as an interaction in the model with WSPT ratings as the outcome variable. The modifying factors CC, E-CLSA, SL PROFICIENCY, and TL PROFICIENCY were added as co-variates in line with the research questions. Since the main effects for CC, E-CLSA, and SL were significant in the model with WSPT ToTs as the outcome variable, interactions between CC and E-CLSA, as well as SL and SL PROFICIENCY were added. The TL is the L1 German and the SLs are the L2 English and the L3 Swedish. The models are presented in Table 29 and Table 30 below.

Table 29

Reverse CLI in German from Swedish and English (WSPT similarity ratings)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.3172	-0.6666	-0.0085	0.6197	3.3809

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	2.0869	1.4446
PARTICIPANT	INTERCEPT	0.6897	0.8305
RESIDUAL		1.4068	1.1861

NUMBER OF OBS: 1499, PARTICIPANT, 8; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	3.328e+00	1.281e+00	2.938e+02	2.597	0.00987	**
TIME	-8.881e-02	5.721e-02	1.420e+03	-1.553	0.12075	
ITEM TYPE	4.969e-01	5.727e-01	7.230e+01	0.868	0.38844	
SOURCE LANGUAGE	-4.234e-01	5.957e-01	8.448e+01	-0.711	0.47915	
TIME ON TASK	4.642e-02	1.804e-01	1.421e+03	0.257	0.79693	
COGNITIVE CONTROL	3.237e-01	3.326e-01	5.863e+00	0.973	0.36881	
E-CLSA PSYCHOTYPOL	-8.568e-02	5.796e-02	1.423e+03	-1.478	0.13956	
SL PROFICIENCY	1.036e-03	3.197e-04	1.421e+03	3.240	0.00122	**
TL PROFICIENCY	-2.948e-04	1.298e-03	4.911e+02	-0.227	0.82038	
ITEM TYPE:TOT	-2.500e-01	2.378e-01	1.420e+03	-1.052	0.29318	
TIME:ITYP:TOT	1.035e-01	8.362e-02	1.419e+03	1.237	0.21617	
TIME:ITYP:SLANG:TOT	4.028e-03	1.121e-01	1.419e+03	0.036	0.97133	

REML criterion at convergence 5068.2

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). SL = L2 English or L3 Swedish. TL = L1 German. INTERCEPT is set to time 1, baseline items, source language English, as well as mean CC, ToTs, psychotypology, SL proficiency, and TL proficiency. TOT and COGNITIVE CONTROL have been normalized using Z-transformation. Formula: rating ~ time * itemtype * source language * scale(ToT) + scale(cc) + E-CLSA + SLP DIALANG + TLP DIALANG + (1|item) + (1|participant).

Table 30

Reverse CLI in German from Swedish and English (WSPT ToTs)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.2762	-0.5634	-0.1278	0.3563	11.5276

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	91335	302.2
PARTICIPANT	INTERCEPT	454425	674.1
RESIDUAL		929050	963.9

NUMBER OF OBS: 1499, PARTICIPANT, 8; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	6007.8329	1108.2267	413.3003	5.421	< 0.001	***
TIME	-128.6418	45.7109	1424.3952	-2.814	0.0050	**
ITEM TYPE	79.8472	208.0893	322.3450	0.384	0.7014	
SOURCE LANGUAGE	712.1505	321.1655	1011.6417	2.217	0.0268	*
COGNITIVE CONTROL	1123.6738	318.8759	10.2662	3.524	0.0053	**
E-CLSA PSYCHOTYPOL	-567.5553	110.3567	1288.9332	-5.143	< 0.001	***
SL PROFICIENCY	-0.4323	0.3473	1429.0227	-1.245	0.2134	
TL PROFICIENCY	-1.7648	1.0742	499.7693	-1.643	0.1010	
TIME:ITEM TYPE	15.8454	62.2177	1423.0522	0.255	0.7990	
TIME:SLANGUAGE	365.9423	77.5970	1429.5205	4.716	< 0.001	***
E-CLSA:CC	-206.2679	43.2960	1358.3691	-4.764	< 0.001	***
LANGUAGE:SL P	-1.6773	0.5319	1419.4683	-3.153	0.0017	**
TIME:ITYP:SLANGUAGE	37.1258	88.0503	1423.0971	0.422	0.6733	

REML criterion at convergence 24829.3

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). SL = L2 English or L3 Swedish. TL = L1 German. INTERCEPT is set to time 1, baseline items, source language English, as well as mean CC, psychotypology, SL proficiency, and TL proficiency. COGNITIVE CONTROL has been normalized using Z-transformation. Formula: ToT ~ time * itemtype * source language + scale(cc) * E-CLSA + SLP DIALANG + language + TLP DIALANG + (1|item) + (1|participant).

As with forward CLI, *reverse CLI* presents itself differently in the two outcome variables. For WSPT similarity ratings, we have a generic learning effect but that is not significant (p = 0.12) as well a statistically significant effect of the source language (p = 0.03) with increased proficiency leading to higher similarity ratings. No item-level effects for similarity evaluations were observed. What comes to ToTs, no item level effects were observed either. We have statistically significant main effects for time (generic learning effect), source language, CC, and psychotypology. These relate to faster ToTs over time, longer ToTs for items with Swedish as the SL (as opposed to English), better CC (lower values better) resulting in faster overall ToTs, and higher perceived similarity across the languages resulting in faster ToTs. There is a significant interaction between time and source language, meaning that the effect of source language (L2 English vs. L3 Swedish) develops differently over time. We also have a significant interaction between CC and psychotypology, SL proficiency, and SL (L2 English vs. L3 Swedish).

5.4.3 Multidirectional CLI in L2 English

In addition to analyzing forward CLI in Swedish and reverse CLI in German, we can look at the L2 English as the TL in which case we have potentially both forward CLI from the L1 German and reverse CLI from the L2 Swedish. Two LMEs were performed: one for WSPT similarity ratings and one for WSPT ToTs as outcome variables. TIME, ITEM TYPE, and SL were added as interactions. TOT was added as an interaction in the model with WSPT similarity ratings as an outcome variable. CC, E-CLSA, SL PROFICIENCY, and TL PROFICIENCY were added as covariates in line with the research questions. The TL is English and the SLs are the L1 German and the L3 Swedish. The models are presented in Table 31 and Table 32 below.

Table 31

Forward and reverse CLI in English from German and Swedish (WSPT similarity ratings)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.8052	-0.6278	0.0080	0.5878	3.2255

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	1.9111	1.3824
PARTICIPANT	INTERCEPT	0.5398	0.7347
RESIDUAL		1.4610	1.2087

NUMBER OF OBS: 1440, PARTICIPANT, 8; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	2.119e+00	7.124e-01	1.995e+02	2.975	0.0033	**
TIME	-4.019e-02	6.279e-02	1.368e+03	-0.640	0.5222	
ITEM TYPE	7.680e-01	5.621e-01	8.069e+01	1.366	0.1757	
SOURCE LANGUAGE	5.242e-02	6.522e-01	1.441e+02	0.080	0.9361	
TIME ON TASK	-1.368e-01	4.128e-02	1.378e+03	-3.313	< 0.001	***
COGNITIVE CONTROL	5.842e-02	3.031e-01	6.312e+00	0.193	0.8531	
E-CLSA PSYCHOTYPOL	2.555e-01	1.039e-01	8.638e+02	2.459	0.0141	*
SL PROFICIENCY	7.158e-05	5.948e-04	1.241e+03	0.120	0.9042	
TL PROFICIENCY	9.973e-04	4.538e-04	1.306e+03	2.198	0.0282	*
TIME:ITEM TYPE	-5.819e-02	8.501e-02	1.365e+03	-0.684	0.4938	
TIME:ITYP:SLANG:TOT	1.947e-04	1.203e-01	1.365e+03	0.002	0.9987	

REML criterion at convergence 4907.7

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). SL = L1 German or L3 Swedish. TL = L2 English. INTERCEPT is set to time 1, baseline items, source language German, as well as mean CC, ToTs, psychotypology, SL proficiency, and TL proficiency. TOT and COGNITIVE CONTROL have been normalized using Z-transformation. Formula: rating ~ time * itemtype * source language * scale(ToT) + scale(cc) + E-CLSA + SLP DIALANG + TLP DIALANG (1|item) + (1|participant).

Table 32

Forward and reverse CLI in English from German and Swedish (WSPT ToTs)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.2197	-0.5570	-0.1359	0.3291	7.6134

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	51499	226.9
PARTICIPANT	INTERCEPT	440180	663.5
RESIDUAL		1167047	1080.3

NUMBER OF OBS: 1440, PARTICIPANT, 8; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	2001.6887	554.9047	145.2136	3.607	< 0.001	***
TIME	-72.2275	56.0858	1369.5482	-1.288	0.1980	
ITEM TYPE	256.2900	235.9851	689.9393	1.086	0.2778	
SOURCE LANGUAGE	-1172.0752	377.0642	1275.1471	-3.108	0.0019	**
E-CLSA PSYCHOTYPOL	320.3285	92.4974	848.8577	3.463	< 0.001	***
COGNITIVE CONTROL	360.4987	273.3299	6.2311	1.319	0.2336	
SL PROFICIENCY	-1.3944	0.5304	1236.2497	-2.629	< 0.01	**
TL PROFICIENCY	1.4341	0.4038	1312.8074	3.551	< 0.001	***
TIME:ITEM TYPE	5.4704	75.9744	1365.9306	0.072	0.9426	
TIME:SLANGUAGE	193.7480	83.2284	1370.9548	2.328	0.0201	*
TIME:ITYP:SLANGUAGE	-97.0411	107.4440	1365.9306	-0.903	0.3666	

REML criterion at convergence 24157.7

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). SL = L1 German or L3 Swedish. TL = L2 English. INTERCEPT is set to time 1, baseline items, source language German, as well as mean CC, psychotypology, SL proficiency, and TL proficiency. COGNITIVE CONTROL has been normalized using Z-transformation. Formula: ToT ~ time * itemtype * source language + scale(cc) + E-CLSA + SLP DIALANG + TLP DIALANG (1|item) + (1|participant).

In terms of results, we can observe different effects in the two outcome variables. No item-level effects were observed in either. There are statistically significant main effects of ToTs, psychotypology, and TL proficiency in WSPT similarity ratings. Longer ToTs lead to lower similarity ratings, higher perceived similarity to higher ratings, and higher TL proficiency to higher similarity ratings. No statistically significant interactions were observed. For ToTs as the outcome variable, there are statistically significant main effects for SL, psychotypology, SL proficiency, and TL proficiency. Also, there is a statistically significant interaction for SL and time. Swedish-derived items have overall faster ToTs than German-derived items but this difference reduces over time. Higher perceived similarity results in longer ToTs. Better CC (smaller score is better) results in faster ToTs.

5.5 Discussion

The present study expanded on the previous studies in this dissertation in that it presented longitudinal data from naturalistic learners and that it investigated all six potential directions of CLI in the MML. The assumption was that multilingual language acquisition results in fluctuating competence. The participants were L1 German speakers of L2 English acquiring L3 Swedish. Their acquisition of the L3 Swedish was tracked for a period of ten months. The focus of the investigation was on two aspects relating to multilingual language acquisition: the *direction* of CLI, and the effect of *modifying factors.* Furthermore, the present study aimed to investigate how these are effected by stage of acquisition.

The first topic of interest in the present study was *direction* of CLI. The point of departure was the assumption that a speaker's languages are not functionally separate (Kroll, Dussias, Bice & Perrotti, 2015). From this perspective, and given the high proficiency in all three languages at the end of the data collection period, the expectation was that some form of CLI would be observed in all six potential directions.

For *forward CLI*, there was a major initial increase in similarity ratings in the L3 Swedish based on translation ambiguity with the L1 German. If the L1 German was deemed to be the most suitable language to act as a mediator between L3 forms and language independent concepts, then it would by default be active. It is also possible, that the participants had not yet acquired proper regulations skills for the L1 German at that point. The ToTs for forward CLI from L1 German to L3 Swedish increased throughout the duration of the study. On the other hand, there seems to be an increasing preference (or lack of inhibition thereof) for the L2 English in L3 processing. However, based on the statistical analysis, comparatively, and despite the increasing trend for the effect of the L2 English, the L1 German is the predominant supplier for meaning-based CLI in the L3 Swedish when measured in ToTs. This would suggest that both the L1 German and the L2 English may result in forward CLI in the L3 Swedish, but that the quantity fluctuates.

The curvilinear trends in forward CLI support the developmental hypotheses presented in the PM. The results indicate fluctuation in performance as well as that the L3 forms are, at least in some way, connected to the L2 and the L1 representations. This finding also aligns with the predictions of the Revised Hierarchical Model (Kroll & Stewart, 1994) if adapted for third language acquisition. There are no indications, however, that the L3 would have reached any form of independence from the L1 and the L2 by the end of the intensive language course (as proposed as the end-state by Jiang's (2000) L1 Lemma Mediation Hypothesis and in a milder form by the PM). It can, however, of course be the case that the learners have not *yet* become proficient enough in their L3 for this to have taken place. With respect to Jiang's (2000) proposal of learners being dependent of conscious rules for resolving translation ambiguity, the relationship between similarity ratings and ToTs in the present study corroborates this hypothesis.

For *reverse CLI*, Schmid & Köpke (2017) have proposed that this type of multilingual language acquisition would lead to attrition in the mother tongue. However, Linck et al. (2009), as well as Opitz (2013), have suggested that any initial hit in the L1 that is related to extreme shifts in language use is temporary. No reverse item-level effects were found suggesting global inhibition of the L3. The quantity of global inhibition in reverse CLI seems to be driven by psychotypology, aptitude, and the combination of the two (alas, the SL – either L2 English or the L3 Swedish – is inhibited when it is perceived to be necessary and when the cognitive resources are available). The hypothesis that the reverse effects are temporary was not corroborated. Most likely, this means that the participants (still being immersed in the L3 environment during the last data collection session) have not *yet* managed to acquire sufficient regulation skills in relation to the extreme increase in their language skills in the L3 Swedish. The language-level effects in the previously acquired languages L1 German and L2 English align with Higby and colleague's (2020) hypothesis that immersed language learners rely on *global* inhibition of the pre-existing language even when using exclusively the L1. Based on the results of the present study, we could expand that hypothesis to the previously acquired L2. A statistically significant effect was found for increased ToTs in both the L1 German and the L2 English for all tested items over time meaning that the participants became progressively slower.

It seems that CLI presents itself in a different manner for forward and reverse CLI. Item-level effects were only found in the forward CLI condition, whereas language-level effects were predominantly found in the reverse condition. These findings align, respectively, with the hypotheses of selective property-wise transfer as proposed by the Scalpel Model by Slabakova (2016) and the wholesale transfer as proposed by the TPM (albeit only for the initial state) by Rothman (2015). Note, though, that neither of these models make specific predictions about vocabulary. The explanation to the observed effects is likely processing economy: it is presumably less costly to inhibit a less proficient language wholesale in reverse CLI than to do so to a more proficient language in forward CLI. There are also likely little to no positive effects of property-wise inhibition patterns in reverse CLI since the L1 has already been acquired to a native(-like) proficiency. On the other hand, the pre-existing languages – particularly given the close resemblance between the three languages German, English, and Swedish in the present study – can be used as a resource in subsequent language acquisition.

The second topic of interest in the present study was the effect of the modifying factors *aptitude* (operationalized as CC) and *psychotypology* (treated as a conscious construct and measured overtly). The three target languages were not affected by CC and psychotypology in a similar way. The L3 Swedish was affected by both the L1 German and L2 English at the item level and the effect not was modulated by CC or psychotypology. This would suggest that forward CLI in the L3 is unavoidable. For the L1 German, reverse CLI (from the L2 English and the L3 Swedish) in the form of higher similarity ratings was not observed and this result was not dependent on aptitude or psychotypology. In processing cost, reverse CLI in the L3 German from both the L2 English and the L3 Swedish was modulated by psychotypology and cognitive control. Higher perceived similarity in the E-CLSA and better CC resulted in faster ToTs. This would suggest that better regulation skills can result in less reverse CLI and that the attempt to regulate later acquired languages depends on perceived similarity between the languages.

In the L2 English, higher perceived similarity in the E-CLSA psychotypology task resulted, unexpectedly, in longer time spent on task. These results were not modulated by CC. That is, a higher perceived global language similarity between the SLs (L1 German and L3 Swedish) and the TL L2 English resulted in longer time on task in the conditions where English is the TL. The presented hypotheses are unable to account for this result (with an estimate of 320ms and $p = < 0.001$), but given that there is a major shift in language use patterns (introduction of a new language, lack of use of the L1), it might represent some form of hypercorrection or avoidance (Schachter, 1974).

5.6 Conclusion

The main purpose of the present study was to observe the development of conceptual-lexical relationships in naturalistic language acquisition from a longitudinal perspective with the point of departure that a speaker's languages are not functionally separate.

Longitudinal data was collected from German (L1) and English (L2) speaking learners of Swedish (L3). Development of all three aforementioned languages was of interest in the present study. A new test for cross-linguistic influence, the Word Pair Similarity Perception Task, was developed for the present study. In the task, the participants evaluated similarity of two words and the relationship of those words was manipulated. Some of the word pairs had translation ambiguity and some did not. The outcome variables were the similarity ratings and the time it took to make the evaluations.

CLI was found to be multidirectional, albeit operating differently for forward and reverse CLI. Forward CLI seemed to take place at the level of individual items and not be dependent of the language or participant level modifying factors. Thus, forward CLI seems unavoidable. On the other hand, reverse CLI seemed to take place at the level of inhibiting the language wholesale and was modulated by available cognitive resources and perceived similarity.

An unexpected effect was found when the participants were tested in the L2 English. That is, when the similarity between the source language (either L1 German or L3 Swedish) and L2 English was perceived higher, the participants spent more time on their evaluations. It will be interesting to see whether the same effect presents itself in the larger sample population in Study 4 (Chapter 6 in this thesis).

6 Cross-Sectional Aspects of Naturalistic L3 Lexical Acquisition

6.1 Introduction

The results of Study 3 (Chapter 5) entail that becoming a multilingual is a multi-faceted process whereby increased vocabulary knowledge also requires the regulatory processes to develop accordingly. Cross-linguistic influence (CLI) was found to be multidirectional. It, however, operated differently depending on the direction. Item-level effects were found for forward CLI, while language-level effects were found for reverse CLI. Furthermore, an unexpected effect of psychotypology was found. Increased perceived similarity between the source languages L1 German and/or L3 Swedish and the target language L2 English resulted in longer time on task in the L2 English representing some form of avoidance (Schachter, 1974). The present study is essentially a replication of Study 3 using a cross-sectional method. The strength of Study 3 is its longitudinal design. However, the sample size is small. Furthermore, it is not possible to distinguish between effects that are temporary in nature (as suggested by Linck, Kroll & Sunderman, 2009; and Opitz, 2013) and that specifically relate to the intensive learning experience from those that constitute attrition. The present study was designed to address these potential shortcomings and with a cross-sectional design, a larger sample size was achieved.

As with Study 3, all six potential directions of CLI were investigated in the present study. There are three languages *in which* CLI can be observed and, in each of the three cases, there are two languages *from which* CLI can be observed. The language in which CLI is observed, i.e., L1 German, L2 English, or L3 Swedish, is referred to as the target language (TL) and the language *from which* CLI is observed, i.e., L1 German, L2 English, or L3 Swedish, is referred to as the source language (SL). All statistical modeling has been performed so that there is always one TL and two potential SLs.

At the onset of third language (L3) acquisition, a bilingual has lexical competitors – i.e., other words that are connected to a particular word either due to similar meaning, similar form, or through being translation equivalents – in a particular language as well between the two languages. We can refer to these as within-language and between-language competitors (Marian & Spivey, 2003). As the proficiency in the L3 increases, additional demands are placed on the regulation of cross-language activation (CLA) in a system that now includes between-language competitors in three languages. This, then, results in "periods of stagnation, re-learning, and attrition of L2/L3 as well as L1 lexis" (Ecke, 2015, p. 154). Based on the findings of Study 3, some form of CLI in all six potential directions is expected.

In Studies 1-3, CLI presented itself differently in outcomes and processing. When CLI takes place due to CLA – that is *spreading activation*, *non-selective access* (see Section 2.1), and/or access routes between the lemma and the lexeme that involve an additional language (as proposed by Kroll & Stewart, 1994, Jiang, 2000, and Hall & Ecke, 2003) – the effect of CLA is expected to be visible in the outcomes. CLA can, however, be regulated through inhibition, or adjusted using conscious rules (cf. Jiang, 2002) depending on available cognitive resources and task demands. This means that we might not be able to observe the effects of CLA in the outcomes, but we should be able to observe resulting CLI in processing (operationalized as ToTs or RTs, depending on task type).

With respect to the method in the present study, data was collected both from learners based in Germany and Sweden. Some participants were more proficient in their L3 than their L2. Also, instead of progress in the intensive language program, the variables *source language (SL) proficiency*, *target language (TL) proficiency*, *age of onset (AO)*, and *length of exposure (LoE)* constitute proficiency variables, and the effect of these is analyzed separately in the statistical modeling.

6.2 Aim and predictions

The main aim of the present study was to observe the development of conceptual-lexical relationships in naturalistic language acquisition from a cross-sectional perspective. Specifically, the present study attempts to untangle the effects of *source language proficiency, target language proficiency*, *age of onset*, *length of exposure*, and *intensity of exposure* on CLI. Given this aim, the following three research questions have been formulated:

RQ1 If present, is cross-linguistic influence (CLI) in the multilingual mental lexicon (MML) unidirectional or multidirectional?

RQ2 To what extent is CLI in the MML affected by psychotypology?

RQ3 What is the effect of source and target language proficiency, age of onset, and length of exposure on the direction of CLI?

The first research question relates to the direction of CLI. All six potential directions were of interest. The second research question relates to the role of psychotypology, which in the present study is treated as a factor with between-participant variation and operationalized as *perceived similarity*. Furthermore, it is measured overtly in that the participants were asked about the perceived similarity using an explicit task. The third research question relates to the effect of proficiency in the three languages and is operationalized by a range of different sub-components that represent variation in the learning experience and (language-wide) learning outcomes.

The first investigated variable in the present study was *direction*. The results in Study 3 provide us with some predictions. In study 3, for forward CLI, the L1 German was the predominant supplier of CLI in the L3, although there was an increasing trend for L2 English-derived CLI with increasing proficiency. Based on the Revised Hierarchical Model (Kroll & Stewart, 1994), the Parasitic Model (Hall & Ecke, 2003), and The L1 Lemma Mediation Hypothesis (Jiang, 2000; 2002), similar results are expected in the present study. For reverse CLI, it is hypothesized that the results in Study 3 might be, to a large extent, a result of the intensive language learning experience in the L3, i.e., progressing from CEFR A1 to C1 in the L3 Swedish in ten months. Since there is a shortage of both models making explicit predictions as well as empirical research about reverse CLI caused by between-language conceptual non-equivalence that is not related to attrition or temporary effects of immersive acquisition, no specific predictions can be made.

The second investigated variable in the present study was *psychotypology*. As noted previously, in Study 3, the L1 German was the primary source of forward CLI in the L3 Swedish with increased influence from the L2 English. These findings were not modulated by psychotypology. For reverse CLI observed in the L1, no effect of psychotypology was observed in the outcomes. Still, there was a statistically significant effect in processing cost in that higher perceived similarity between the SL (L2 English or L3 Swedish) and the TL (L3 German) led to faster responses. The results were expected. However, unexpected effects of psychotypology were found when the participants were tested in the L2 English. Increased perceived similarity between the SL (L1 German or L3 Swedish) and TL (L2 English) led to longer time on task in the TL. In the case this was due to the small sample size, we should not observe the same effects in the present study. However, if the results represent some form of hypercorrection or avoidance (Schachter, 1974), the effect should be observed in the present study as well. In line with the findings in Study 3, we would expect to find no effect of psychotypology in forward CLI if the participants depend on resolving translation ambiguity using item-specific conscious rules. Also, in line with Study 3, we would expect to find an effect of psychotypology in language-level inhibition of the later acquired languages. However, with respect to the finding in Study 3, that higher perceived similarity has an adverse effect on processing in the L2 English, no specific predictions can be made.

The third investigated variable in the present study was *proficiency*. In Study 3, SL and TL proficiency, AO, and LoE had relatively little variation between the participants. As a group, an effect of TL (L3 Swedish) proficiency in forward CLI outcomes was observed, which is expected given that the participants progressed from relative beginners to a very high proficiency during the longitudinal data collection. We can expect similar results (cf., RQ3) in the present study, either in terms of measured SL and TL proficiency, or through AO and LoE effects. Furthermore, in Study 3 for reverse CLI measured in the L1 German, proficiency in the SLs (L2 English and L3 Swedish) affected outcomes but not processing in Study 3. We can expect that there is a difference in the present study depending on the location of the participants, with participants residing in Sweden being more affected due to L1 attrition having presumably fewer opportunities to use the L1 German. Finally, in Study 3, proficiency in the TL (L2 English) predicted CLI in outcomes. Proficiency in both the TL and SLs predicted CLI in processing.

We can expect similar results in Study 4 since the L2 English was not subject to the intensive acquisition in Study 3.

6.3 Method

6.3.1 Participants

The participants were native speakers of L1 German, spoke fluent L2 English (CEFR B2 or higher, but not native), and were learning L3 Swedish (CEFR A2 or higher, but not native). The distinction between the L2 and the L3 was done based on age of acquisition and in the case of some of the participants does not represent proficiency. The participants fell under two distinct categories: they either lived in southern Sweden and had acquired Swedish predominantly in Sweden, or they were studying Swedish in Germany either at University of Greifswald or at University of Konstanz. Many of the participants in the latter category had resided in Sweden at some point in their lives. Data collection took place in Lund, Malmö, Greifswald, and Konstanz between 2017 and 2020.

There were a total of 29 participants, ranging from 19 to 47 in age (M = 27.1, SD = 6.38). Most of the participants were female (69%). All participants had a minimum of high-school education completed. AO for German was zero with on average 26.4 years of self-reported exposure (SD = 8.03, Min-Max 19-47), mean AO for English was 8.59 (SD = 3.01, Min-Max = 3-14) with on average 17 years of self-reported exposure (SD = 5.00, Min-Max 8-34), and mean AO for Swedish was 18.6 (SD = 7.93, Min-Max = 3-40) with on average 7.36 years of exposure (SD = 5.57, Min-Max = 0.5-21).

On average, the participants' most proficient language was the L1 German with an average score of 835[58] (SD = 118, Min-Max = 495-1,000) in the DIALANG lexical placement module. The second most proficient language was the L2 English with an average score of 818 (SD = 139, Min-Max = 482-980). The least proficient language, and also the one with the most variation in the sample, was the L3 Swedish with an average score of 751 (SD = 214, Min-Max = 139-941). Thirteen (44.8%) participants reported being based in Sweden and sixteen

[58] The DIALANG scores range from 0-1000. A score of 901-1000 represents a native-speaker range. Scores of above 601 represent vocabulary sizes of advanced language learners.

(55.2%) in Germany. Twelve (41.4%) of the participants had higher proficiency scores in their L3 Swedish than their L2 English. Seven of these participants were based in Sweden at the time of data collection. Based on self-evaluations of their skills in their German, none of the participants reported major effects of L1 attrition, with all participants rating their reading, writing, listening, and speaking skills at seven on a seven-point Likert-scale.

6.3.2 Instruments

No new instruments were developed for the purposes of Study 4. The used experimental instruments – which are visualized in Figure 35 below – are the Explicit Cross-Linguistic Similarity Assessment (E-CLSA) psychotypology task (see 4.3.3.3), the lexical placement module from DIALANG (see 5.3.2.2), and the Word Pair Similarity Perception Task (WSPT) to measure CLI (see 5.3.2.3).

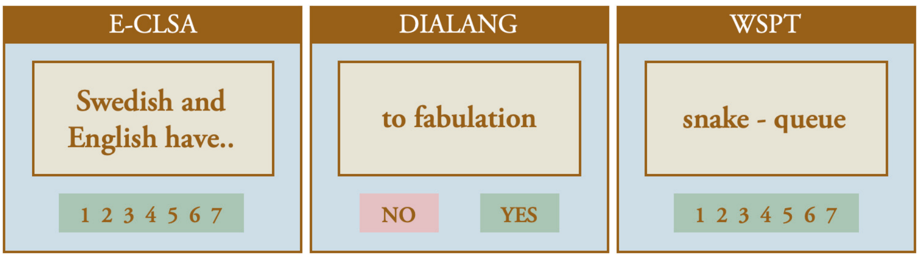


Figure 35. Experimental instruments in Study 4

Since compensating participants in salary (like in Study 3) was not possible, attempts were made to limit the length of the data collection session. In addition to streamlining the procedure and limiting psychotypology assessment to explicit evaluations of psychotypology, WM and CC were dropped. After these adjustments, participation took approximately 90 minutes. This means that the only aptitude measure collected was self-evaluation of language learning aptitude, which has moderate correlations with more objective aptitude measures (see Table E6 in Appendix E).

6.3.3 Procedure

The data collection took place in one sitting (visualized in Figure 36 below). The non-language specific instruments, including the background questionnaire (available in Appendix J), were administered first. The language-specific blocks then followed. Each of the three language-specific blocks included only the target language, meaning that everything that was presented to the participants was done in the language of that particular block, including all instructions. Each block started with a short film of approximately five minutes in the TL to facilitate the participant's shift of language mode. This was followed by a two-minute oral production task where the participant was asked to either describe the preceding film, another film, or a book of their choosing. After these, the lexical placement module from DIALANG was administered disguised as a lexical decision task. Finally, the participants took two conditions in the WSPT word pair similarity task, each with the target language of the particular block and with the two other languages as the source languages. The participants were not informed about the translation ambiguity across languages. Debriefing was done at the end, either in Swedish or English depending on the participant's preference.

BACKGROUND ENGLISH	Welcome Consent for participation[59] Background questionnaire E-CLSA (psychotypology) Break
BLOCK 1 ENGLISH	The participants watched a video in TL (5min) Speech task in TL (2min) DIALANG WSPT Similarity evaluations (15 treatment and 15 baseline items from two conditions with the same target language and 60 fillers – a total of 120 items, break after 60 word pairs) Break
BLOCK 2 GERMAN	The participants watched a video in TL (5min) Speech task in TL (2min) DIALANG WSPT Similarity evaluations (15 treatment and 15 baseline items from two conditions with the same target language and 60 fillers – a total of 120 items, break after 60 word pairs) Break
BLOCK 3 SWEDISH	The participants watched a video in TL (5min) Speech task in TL (2min) DIALANG WSPT Similarity evaluations (15 treatment and 15 baseline items from two conditions with the same target language and 60 fillers – a total of 120 items, break after 60 word pairs) Post-experiment interview / debriefing (Swedish or English)

Figure 36. Procedure in Study 4

[59] A privacy policy and data management description was provided to the participants. In accordance with GDPR, the European Union General Data Protection Regulation (2016/679), the participants consented in writing to the following: 1) that their personal information will be stored separately from research data, 2) that personal information will be stored in relation to any compensation in accordance with the Swedish accounting legislation (BFL 1999:1078) by the university centrally, 3) that the collected but anonymized research data will be publicly available and may be used for further research purposes without additional consent from the participant, 4) and that it is possible to retract one's participation up until the point of publication. None of the conditions by the Swedish Research Council that mandate ethical review apply for the present study. The participants were compensated with movie tickets.

6.3.4 Data analysis, transformations, and structure

Two cases of obvious erroneous key presses with values of 1ms were removed from the WSPT time on task (ToT) data set. No further lower spectrum trimming was deemed necessary as the rest of the values (all 600ms or higher) were found to be within normal range. Fifty cases of obvious unplanned breaks (ToT ≥ 10,000ms) representing 0.48% of the data were removed. The translation ambiguity condition was more effected than the baseline condition (0.70% vs. 0.35%). The WSPT tasks were not time-restricted, as the task was designed to allow introspection (see Section 2.3.1), and thus no (additional) strict upper spectrum trimming was done. In general, given the possibility for introspection, the values should be considered as values of time on task rather than response times per se.

Global measures of psychotypology from the E-CLSA were computed from the area specific questions by giving equal weight for each question. Three separate estimations of psychotypology resulted, each representing perceived similarity between two of the participant's languages. These have been tagged to responses in the WSPT data based on the particular item's SL/TL pairing. Similarly, proficiency scores: SL proficiency, SL AO, and SL LoE, as well as TL proficiency, TL AO and TL LoE were tagged to each response in the WSPT data.

All statistical analysis was performed in R version 3.6.0 (R Core Team, 2018) using R Studio version 1.2.1335 as the graphical user interface. Data exploration was done using Jamovi version 1.0.8.0 (The Jamovi Project, 2019). Plotting was performed using sjPlot version 2.8.5 and then adjusted to the house style.

There are six outcome variables in Study 4: WSPT similarity ratings and time on task (ToT) in each of the three target languages L1 German, L2 English, and L3 Swedish representing forward and reverse CLI. Separate linear mixed effects models were performed for each learning outcome representing outcomes and processing in each of the target languages. For all models, random intercepts were included for PARTICIPANT and ITEM. The main (fixed) factors ITEM TYPE and SOURCE LANGUAGE were added as interactions. LOCATION of residency (Germany, Sweden), E-CLSA psychotypology evaluations, SL PROFICIENCY, AO, and LOE, as well as TL PROFICIENCY, AO, and LOE were added as co-variates. In the models with WSPT ratings as outcome variables, TOT was added as a co-variate as well. Furthermore, in the models where German was the target language, the co-variate TL AO was removed since there was no variation – all participants reported L1 AOs of zero years (this was a selection criteria for participation in the experiment). The co-variates TOT, E-CLSA, SL PROFICIENCY, TL PROFICIENCY, AO, and LOE were normalized using Z-transformation. Marginal effects (rescaled to the original scale) from the linear mixed effects models were extracted for the categorical variables SOURCE LANGUAGE and ITEM TYPE for illustrative purposes in Figures 37, 38, and 39.

6.4 Results

6.4.1 Forward CLI in L3 Swedish

Forward CLI to the L3 Swedish from the L1 German and the L2 English seem to operate differently. Figure 37 below indicates that CLI can predominantly be observed from the L1 German in ToT time course data while CLI from the L2 English presents itself in the WSPT ratings. In the linear mixed effects models presented in Tables 33 and 34 below, ToTs have a significant effect on WSPT ratings (Estimate = 0.08 points, p = 0.05). Item type effects are not significant in either ratings or ToTs, but a main effect of estimated 255ms is nearing significance for ToTs (p = 0.09).

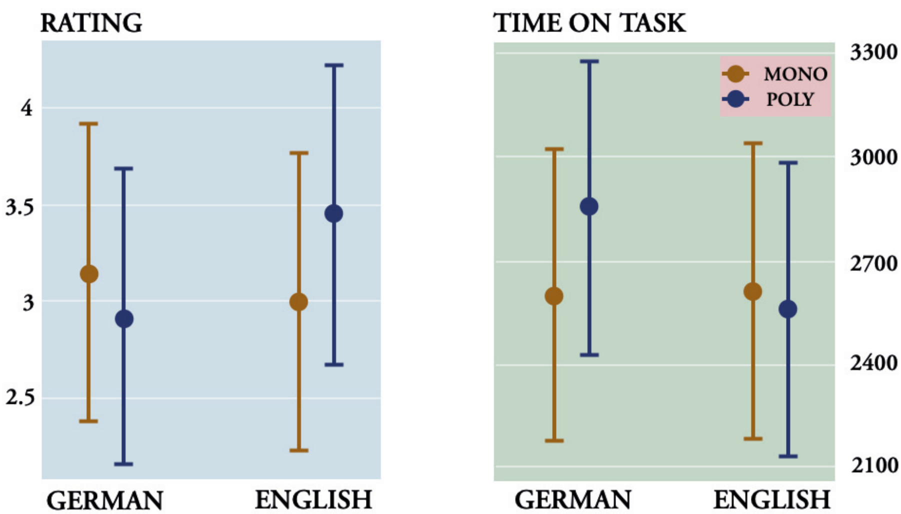


Figure 37. Marginal effects of WSPT ratings in Swedish (Likert 1-7 and ToTs in ms) by source language and item type. MONO = no translation ambiguity and POLY = with translation ambiguity.

Neither of the two potential SLs is a privileged source of CLI and the quantity of forward CLI in the L3 would be modulated by psychotypology, AO, or LoE.

Table 33

Forward Cross-Linguistic Influence in L3 Swedish (WSPT ratings)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.4814	-0.6499	-0.0652	0.6276	3.0864

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	1.4456	1.2023
PARTICIPANT	INTERCEPT	0.5717	0.7561
RESIDUAL		1.9594	1.3998

NUMBER OF OBS: 1678, PARTICIPANT, 28; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	3.14337	0.38822	89.47642	8.097	2.66e-12	***
ITEM TYPE	-0.22934	0.44963	56.03495	-0.510	0.612	
SOURCE LANGUAGE	-0.14517	0.50374	88.04486	-0.288	0.774	
LOCATION (SE)	0.10569	0.31458	23.77462	0.336	0.740	
TIME ON TASK	0.08016	0.04090	1625.92512	1.960	0.050	*
E-CLSA PSYCHOTYPOL	0.10255	0.07367	1120.70110	1.392	0.164	
SL PROFICIENCY	0.04034	0.07083	1375.87626	0.569	0.569	
SL AGE OF ONSET	0.14070	0.12451	1178.40683	1.130	0.259	
SL LENGTH OF E	0.06581	0.10284	1162.96516	0.640	0.522	
TL PROFICIENCY	0.03963	0.16763	25.58000	0.236	0.815	
TL AGE OF ONSET	0.26006	0.17870	28.43394	1.455	0.157	
TL LENGTH OF E	0.11692	0.18145	27.32368	0.644	0.525	
ITYP:SLANGUAGE	0.68024	0.63582	56.01942	1.070	0.289	

REML criterion at convergence 6165.7

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to baseline items, German as SL, and Germany as location. ToTs, E-CLSA scores, DIALANG scores, AOs, and LoEs have been normalized using Z-transformation. Formula: rating ~ source language * item type + scale(ToT) + location + scale(E-CLSA) + scale(SL proficiency) + scale(SL AO) + scale(SL LoE) + scale(TL proficiency) + scale(TL AO) + scale(TL LoE) + (1|participant) + (1|item).

Table 34

Forward Cross-Linguistic Influence in L3 Swedish (WSPT time on task)

Scaled residuals

	MIN	1Q	MEDIAN	3Q	MAX
	-2.2374	-0.5954	-0.1621	0.3847	5.3363

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	118868	344.8
PARTICIPANT	INTERCEPT	413034	642.7
RESIDUAL		1269888	1126.9

NUMBER OF OBS: 1678, PARTICIPANT, 28; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	2599.59	214.34	54.01	12.128	<2e-16	***
SOURCE LANGUAGE	8.76	235.81	329.38	0.037	0.970	
ITEM TYPE	254.90	148.02	56.01	1.722	0.091	.
LOCATION (SE)	210.96	266.22	23.46	0.792	0.436	
E-CLSA PSYCHOTYPOL	-10.09	59.58	1191.36	-0.169	0.866	
SL PROFICIENCY	89.28	57.16	1419.31	1.562	0.119	
SL AGE OF ONSET	74.73	100.64	1237.09	0.743	0.458	
SL LENGTH OF E	-50.97	83.14	1231.23	-0.613	0.540	
TL PROFICIENCY	-198.57	141.56	25.04	-1.403	0.173	
TL AGE OF ONSET	75.26	150.62	27.71	0.50	0.621	
TL LENGTH OF E	172.50	153.02	26.67	1.127	0.270	
ITYP:SLANGUAGE	-309.20	209.30	55.97	-1.477	0.145	

REML criterion at convergence 28358.4

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to baseline items, German as SL, and Germany as location. E-CLSA scores, DIALANG scores, AOs, and LoEs have been normalized using Z-transformation. Formula: ToT ~ source language * item type + location + scale(E-CLSA) + scale(SL proficiency) + scale(SL AO) + scale(SL LoE) + scale(TL proficiency) + scale(TL AO) + scale(TL LoE) + (1|participant) + (1|item).

6.4.2 Multidirectional CLI in L2 English

Just as in forward CLI in L3 Swedish, longer ToTs result in lower WSPT ratings in L2 English as well, presumably indicating introspection (p = 0.05), i.e., using a conscious rule to adjust one's use of a particular word. The L1 German seems to be the privileged source of CLI in that there is more overt CLI from German and more attempt to inhibit the L3 Swedish as shown in Tables 35 and 36 below. These effects, however, are not significant. CLI in L2 English seems to be modulated by psychotypology, with higher perceived similarity resulting in longer ToTs (Estimate = 170ms, p = 0.04). There are significant effects of earlier SL AO (p = 0.05) and longer SL LoE (p = 0.02) resulting in higher WSPT ratings and higher TL proficiency in lower ratings (p = 0.01). Furthermore, earlier SL AO seems to result in higher processing cost (Estimate = -599ms, p = < 0.01) and earlier TL AO in reduced processing cost (Estimate = 326ms, p = 0.09).

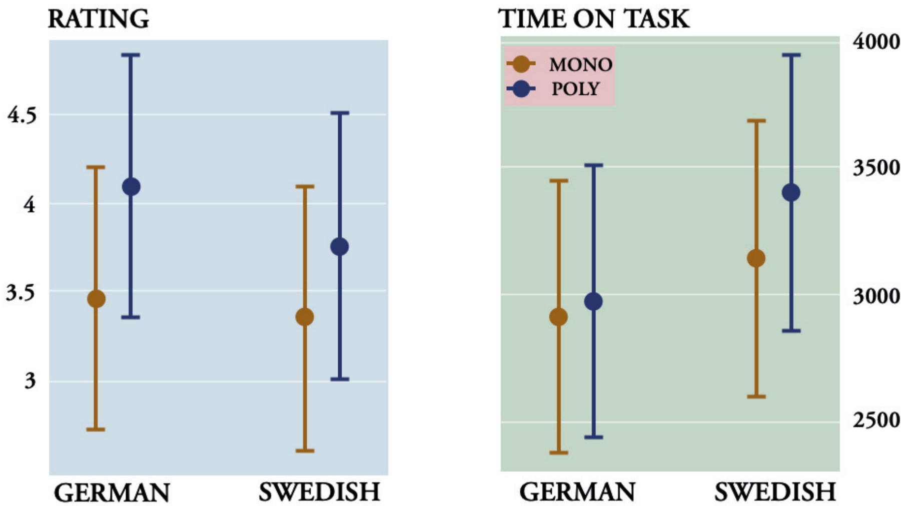


Figure 38. Marginal effects of WSPT ratings in English (Likert 1-7 and ToTs in ms) by source language and item type. MONO = no translation ambiguity and POLY = with translation ambiguity.

Table 35

Forward and Reverse Cross-Linguistic Influence in L2 English (WSPT ratings)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.9207	-0.6336	-0.0285	0.6136	3.3865

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	1.5276	1.2360
PARTICIPANT	INTERCEPT	0.4358	0.6602
RESIDUAL		2.0512	1.4322
NUMBER OF OBS: 1722, PARTICIPANT, 29; ITEM, 60			

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	3.459e+00	3.771e-01	8.331e+01	9.174	2.84e-14	***
SOURCE LANGUAGE	-1.077e-01	4.950e-01	7.390e+01	-0.218	0.8284	
ITEM TYPE	6.387e-01	4.618e-01	5.604e+01	1.383	0.1722	
TIME ON TASK	-8.143e-02	4.186e-02	1.653e+03	-1.945	0.0519	.
LOCATION (SE)	-1.606e-01	2.618e-01	2.339e+01	-0.613	0.5456	
E-CLSA PSYCHOTYPOL	6.651e-02	7.904e-02	3.997e+02	0.841	0.4006	
SL PROFICIENCY	4.336e-02	7.003e-02	6.472e+02	0.619	0.5360	
SL AGE OF ONSET	2.989e-01	1.508e-01	3.906e+02	1.981	0.0482	*
SL LENGTH OF E	3.697e-01	1.600e-01	3.690e+02	2.310	0.0214	*
TL PROFICIENCY	-3.999e-01	1.445e-01	2.591e+01	-2.767	0.0103	*
TL AGE OF ONSET	6.883e-02	1.394e-01	2.542e+01	0.494	0.6257	
TL LENGTH OF E	-3.234e-03	1.487e-01	3.289e+01	-0.022	0.9828	
ITYP:SLANGUAGE	-2.304e-01	6.530e-01	5.601e+01	-0.353	0.7255	

REML criterion at convergence 6395.7

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to baseline items, German as SL, and Germany as location. ToTs, E-CLSA scores, DIALANG scores, AOs, and LoEs have been normalized using Z-transformation. Formula: rating ~ source language * item type + scale(ToT) + location + scale(E-CLSA) + scale(SL proficiency) + scale(SL AO) + scale(SL LoE) + scale(TL proficiency) + scale(TL AO) + scale(TL LoE) + (1|participant) + (1|item).

Table 36

Forward and Reverse Cross-Linguistic Influence in L2 English (WSPT time on task)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.8144	-0.6316	-0.1771	0.4105	4.4322

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	146006	382.1
PARTICIPANT	INTERCEPT	818500	904.7
RESIDUAL		2063167	1436.4

NUMBER OF OBS: 1722, PARTICIPANT, 29; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	2913.38	273.20	41.77	10.664	1.7e-13	***
SOURCE LANGUAGE	234.92	249.07	243.95	0.943	0.34651	
ITEM TYPE	63.98	170.63	56.21	0.375	0.70909	
LOCATION (SE)	467.96	352.05	23.08	1.329	0.19675	
E-CLSA PSYCHOTYPOL	169.59	83.42	793.66	2.033	0.04238	*
SL PROFICIENCY	-89.96	72.95	1040.25	-1.233	0.21778	
SL AGE OF ONSET	-598.66	159.02	723.41	-3.765	0.00018	***
SL LENGTH OF E	-223.33	169.69	696.78	-1.316	0.18858	
TL PROFICIENCY	-40.54	192.18	24.63	-0.211	0.83468	
TL AGE OF ONSET	325.68	185.36	24.53	1.757	0.09138	.
TL LENGTH OF E	275.16	192.03	29.64	1.433	0.16235	
ITYP:SLANGUAGE	194.57	241.07	55.99	0.807	0.42301	

REML criterion at convergence 29928.5

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to baseline items, German as SL, and Germany as location. E-CLSA scores, DIALANG scores, AOs, and LoEs have been normalized using Z-transformation. Formula: ToT ~ source language * item type + location + scale(E-CLSA) + scale(SL proficiency) + scale(SL AO) + scale(SL LoE) + scale(TL proficiency) + scale(TL AO) + scale(TL LoE) + (1|participant) + (1|item).

6.4.3 Reverse CLI in L1 German

Translation ambiguity in the L3 Swedish results in more reverse CLI in both ratings and ToTs in the L1 German, indicating lack of attempt to regulate CLI from the L3 Swedish (see Tables 37 and 38, as well as Figure 39 below). Regulation of CLI from the L2 English would appear to be at a global rather than item level. None of these effects are significant in the LMEs. There are no statistically significant indications that psychotypology, proficiency, AOs, or LoEs would modulate the quantity of CLI. However, there are (not significant) estimations that earlier AO and longer LoE in the SL result in more CLI in both ratings and ToTs. Notably, those participants that are residents of Sweden are estimated to be 517ms (p = 0.06) slower than those participants residing in Germany (which would indicate attrition even when controlled for proficiency, AO, and LoE) in a task that, on the surface level, is performed completely in their L1 German.

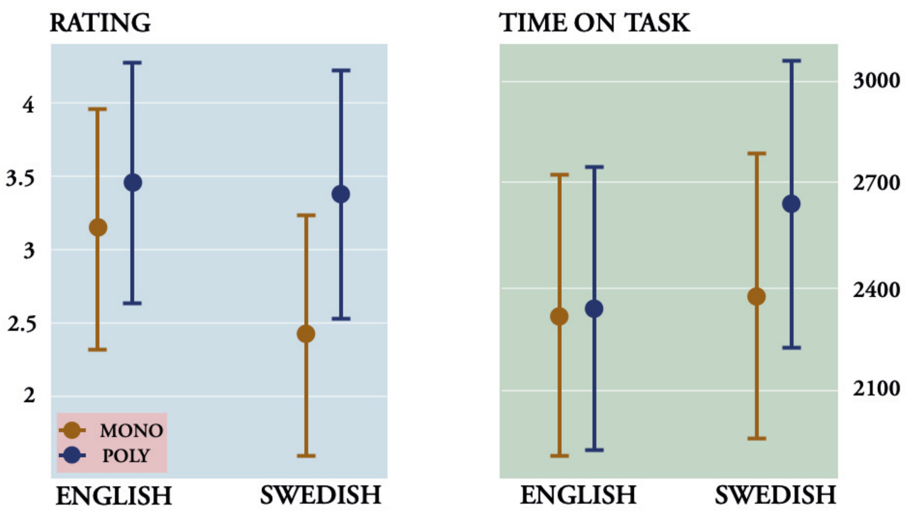


Figure 39. Marginal effects of WSPT ratings in German (Likert 1-7 and ToTs in ms) by source language and item type. MONO = no translation ambiguity and POLY = with translation ambiguity.

Table 37

Reverse Cross-Linguistic Influence in L1 German (WSPT ratings)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-3.1296	-0.6359	-0.0521	0.6065	3.8435

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	1.937	1.3918
PARTICIPANT	INTERCEPT	0.578	0.7602
RESIDUAL		1.523	1.2342

NUMBER OF OBS: 1703, PARTICIPANT, 29; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	3.141e+00	4.198e-01	8.001e+01	7.482	8.28e-11	***
SOURCE LANGUAGE	-7.424e-01	5.315e-01	6.232e+01	-1.397	0.167	
ITEM TYPE	3.150e-01	5.151e-01	5.500e+01	0.612	0.543	
TIME ON TASK	-1.612e-02	3.597e-02	1.644e+03	-0.448	0.654	
LOCATION (SE)	-1.130e-01	3.161e-01	2.525e+01	-0.358	0.724	
E-CLSA PSYCHOTYPOL	2.085e-03	8.542e-02	9.656e+02	0.024	0.981	
SL PROFICIENCY	1.210e-02	6.027e-02	1.288e+03	0.201	0.841	
SL AGE OF ONSET	4.809e-02	8.554e-02	1.477e+03	0.562	0.574	
SL LENGTH OF E	7.339e-02	1.125e-01	1.022e+03	0.653	0.514	
TL PROFICIENCY	-1.674e-02	1.594e-01	2.661e+01	-0.105	0.917	
TL LENGTH OF E	1.525e-01	1.627e-01	3.327e+01	0.937	0.355	
ITYP:SLANGUAGE	6.464e-01	7.349e-01	5.500e+01	0.879	0.383	

REML criterion at convergence 5863.3

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to baseline items, English as SL, and Germany as location. ToTs, E-CLSA scores, DIALANG scores, AOs, and LoEs have been normalized using Z-transformation. Formula: rating ~ source language * item type + scale(ToT) + location + scale(E-CLSA) + scale(SL proficiency) + scale(SL AO) + scale(SL LoE) + scale(TL proficiency) + scale(TL LoE) + (1|participant) + (1|item).

Table 38

Reverse Cross-Linguistic Influence in L1 German (WSPT time on task)

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-1.9877	-0.6121	-0.1820	0.3130	5.5457

Random effects

Groups	Name	Variance	SD
ITEM	INTERCEPT	124005	352.1
PARTICIPANT	INTERCEPT	403836	635.5
RESIDUAL		1516686	1231.5

NUMBER OF OBS: 1703, PARTICIPANT, 29; ITEM, 60

Fixed effects

	Estimate	SE	df	t	*p*	
INTERCEPT	2317.80	208.25	48.29	11.130	6.25e-15	***
SOURCE LANGUAGE	56.56	200.11	154.92	0.283	0.7778	
ITEM TYPE	20.60	153.49	55.12	0.134	0.8937	
LOCATION (SE)	516.78	266.90	24.30	1.936	0.0646	.
E-CLSA PSYCHOTYPOL	36.43	83.32	687.27	0.437	0.6621	
SL PROFICIENCY	-60.91	59.20	1038.66	-1.029	0.3038	
SL AGE OF ONSET	67.43	84.42	1313.97	0.799	0.4245	
SL LENGTH OF E	164.17	109.77	734.51	1.496	0.1352	
TL PROFICIENCY	-21.80	135.38	26.14	-0.161	0.8733	
TL LENGTH OF E	57.00	140.85	34.53	0.405	0.6882	
ITYP:SLANGUAGE	245.44	218.92	55.04	1.121	0.2671	

REML criterion at convergence 29086.7

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to baseline items, English as SL, and Germany as location. E-CLSA scores, DIALANG scores, AOs, and LoEs have been normalized using Z-transformation. Formula: ToT ~ source language * item type + location + scale(E-CLSA) + scale(SL proficiency) + scale(SL AO) + scale(SL LoE) + scale(TL proficiency) + scale(TL LoE) + (1|participant) + (1|item).

6.5 Discussion

The present study is essentially a replication of Study 3 using a cross-sectional method. The focus of the investigation was on three aspects relating to multilingual language acquisition: the *direction* of CLI, the effect of *psychotypology*, and the effect of *proficiency*. This study departed from the assumption that multilingual language learning comes with fluctuating competence and that some of the aspects of the learning process have not been captured by the previous studies in this dissertation. It was assumed that by adopting a cross-sectional design and collecting data from both learners that are immersed in the L3 environment and those who remain in their L1 environment, it should be possible to pinpoint which of the results in Study 3 are specific to the extremely intensive language learning experience and immersion.

The first aspect of interest was *directionality*. In the present study, CLI was found to be multidirectional, in that there were clear results of forward CLI from the L1 German and the L2 English observed in the L3 Swedish, as well as reverse CLI from the L3 Swedish on the L2 English. No results of reverse CLI from the L3 Swedish or the L2 English were found in the L1 German that cannot be explained by L1 attrition.

Recall that statistically significant effects of *reverse CLI in the L1* were found in Study 3 in the intensive language learners. These effects were modulated by cognitive control, psychotypology, and SL proficiency. The privileged source of CLI in the L1 German was the L3 Swedish. These results were *not* replicated in the present study. The finding that such effects could be found in the intensive immersed learners in Study 3, but not in the learners in the present study whose L3 learning experience is characterized by lower intensity and longer LoE, aligns with the suggestion of an initial – and *temporary* – hit in the L1 (Linck, Kroll & Sunderman, 2009; Opitz, 2013). Hence, the effect in Study 3 is presumably specific to immersive language acquisition and separate from attrition. That is, there is a *global inhibition effect* of the L1 that persists even in situations were the L1 is used. We have, in Gyllstad and Suhonen (2017), proposed that this is an initial learning strategy in immersion. In the present study, instead, indications of *attrition* were found in the population that resided in Sweden. The participants in Sweden were estimated to be 517ms (p = 0.06) slower (see Table 38) at responding to tasks in their L1 than the participants who were based in their native Germany.

The second aspect of interest was *psychotypology*. In Study 3, unexpected effects of psychotypology were found for the L2 English, where increased perceived similarity between the SL (either L1 German or L3 Swedish) and the TL (L2 English) led to longer time on task in the TL. It was speculated that this might be due to the small sample size but that if the same results were found in Study 4, the effects would most likely represent some form of hypercorrection or *avoidance* (cf., Schachter, 1974). The same results were found in the present study. The participants were slower at responding to tasks in the TL (L2 English) when they perceived the SL and the TL to be more similar. While the data cannot corroborate the hypothesis of avoidance as the only potential explanation, since the experiment was not specifically designed for that, the likely explanation is that the participants were aware of the occasional translation ambiguity and were particularly careful in the L2 English – a language that was neither their L1 nor a language they were studying at the time of testing.

The third aspect of interest was *proficiency*. What was meant to be the main strength of the present study over Study 3 – namely between-participant variation in different components of the language learning experience – turned out to be its main weakness as well. In the longitudinal design, the participants acted as their own baselines. Furthermore, there was very little variation in the learning environment and experience in Study 3. In the present study, there is a lot of variation in both the learning experience as well as language use. For example, the scores in the DIALANG lexical placement test for the L1 German ranged from 495 to 1,000 (M = 835), with 1,000 representing a perfect score and scores above 901 typical native speaker range. The first assumption would be to see whether the lowest-scoring participants are also those with the longest length of residence in Sweden, which was not the case. The participant with the lowest score in L1 German also had one of the lowest scores in both L2 English, and L3 Swedish. The length of residence for the particular participant in Sweden was five years. In general, LoE in Swedish has a moderate correlation with proficiency in Swedish (r = 0.36) and a rather low correlation with proficiency in German (r = 0.14). Instead, with respect to correlations, a large vocabulary in the L3 Swedish was the best predictor of a large vocabulary in the L1 German (r = 0.59). This result would support the hypothesis of *indirect frequency effects* where repeated parallel co-activation of translation equivalents raise the resting levels of both the target and its translation equivalents (Higby et al., 2020)

6.6 Conclusion

The present study investigated three aspects relating to multilingual language acquisition: the *direction* of CLI, the effect of *psychotypology*, and the effect of *proficiency*. The participants were speakers of L1 German, L2 English, and L3 Swedish. CLI was found to be multidirectional. There were statistically significant results of forward CLI from the L1 and the L2 to the L3, as well as reverse CLI from the L3 to the L2. For psychotypology, the same unexpected result for the L2 English as in Study 3 – where increased perceived similarity between the SL (either L1 German or L3 Swedish) and the TL (L2 English) led to longer time on task in the TL – was found in the present study as well. It was hypothesized that this was a result of *avoidance* (Schachter, 1974). The results for proficiency were inconclusive. It seems that the almost four-fold increase in participants is unable to compensate for the added variation in the background variables in the linear mixed effects models.

Moreover, while the design of the present study allows for insight into when CLI surfaces due to *lack of regulation*, *despite regulation*, or *due to lack of introspection*, future research should attempt to disentangle the effect of regulation that takes place due to *global inhibition* of a particular language and one that takes place due to *introspection* at the level of a particular item. In the present design, both show up as increased time on task and any attempt to disentangle such effects relies on comparing different item types. One potential way to overcome this issue would be to manipulate available cognitive resources by varying enforced time limits. The present study could also be replicated with other non-intensively immersed language learners but with a longitudinal design.

7 Discussion

7.1 General remarks

The empirical work in this thesis has investigated *cross-linguistic influence* (CLI) in the *multilingual mental lexicon* (MML) in second and third language acquisition. The focus has been on how CLI – modulated by *proficiency*, *aptitude*, and *psychotypology* – varies during different stages of learning, stretching from the very initial state of acquisition to a high level of proficiency (CEFR C1 or higher).

This thesis has focused on *meaning-based* CLI, operationalized as *conscious* and *unconscious* lexical activity (Jarema & Libben, 2007) as well as outcomes in *accuracy* and *processing*. The character of interaction between languages has been suggested to change as a result of development in that formal effects (such as *switching* to another language and *inferencing* based on pre-existing languages) dominate at the early stages of second and third language acquisition, while meaning-based effects (such as *semantic extensions*, assuming unambiguous translations) are more prevalent in advanced speakers (Bardel, 2015; Ringbom, 2007). The particular type of meaning-based effect that has been investigated in this thesis is *translation ambiguity* (Eddington & Tokowicz, 2013). Research in this area has focused on speakers with two languages and the present thesis extends this research into third language acquisition.

Three general research questions were formulated to account for the over-arching aims of this dissertation:

RQ1 Is cross-linguistic influence (if present) in the multilingual mental lexicon unidirectional or multidirectional?

RQ2 To what extent is cross-linguistic influence in the multilingual mental lexicon affected by proficiency, aptitude, and psychotypology?

RQ3 Are the aforementioned aspects of directionality and the effect of modifying factors dependent on the stage or type of acquisition?

Four models of multilingual lexical development have been used as a starting point for investigations on directionality in the MML. The Revised Hierarchical Model, RHM, (Kroll & Stewart, 1994) makes testable predictions about CLI from the L1 to the L2 but does not take translation ambiguity into account. The L1 Lemma Mediation Hypothesis (Jiang, 2000; 2002) also makes testable predictions about CLI from the L1 to the L2, but it suggests that learners resolve translation ambiguity using conscious rules. The Revised Hierarchical Model of Translation Ambiguity, RMH-TA, (Eddington & Tokowicz, 2014) makes testable predictions about CLI from the L1 to the L2 with respect to the type of translation ambiguity (form vs. meaning). These three models cover one of the potential six directions of CLI in third language acquisition. Finally, the Parasitic Model (PM) by Hall and Ecke (2003) makes testable predictions about CLI from the L1 to the L3 and from the L2 to the L3 in third language acquisition at the early stages. The three first models propose somewhat linear progression whereas the PM proposes fluctuation during learning. Thus, the PM is useful for classifying instances of temporary lapses in production at the later stages, but it is difficult to pinpoint specific testable hypothesis beyond the initial stages. Finally, none of the four models make any predictions about the remaining three potential directions of CLI in the MML: L3 to L2, L3 to L1, and L2 to L1. There does not seem to be any comparable developmental models of CLI in the MML that would make explicit predictions about the remaining three potential directions of CLI in third language acquisition. With respect to the remaining three, the empirical investigations in this thesis have sailed on relatively uncharted waters.

Four studies were conducted to investigate the aforementioned research questions. Studies 1 and 2 investigated the very initial stages of L2 and L3 learning, respectively, representing data from initial state learners. In Study 1, the L1 was English, and in Study 2, the L1 was Swedish and the L2 was English. These two studies were more experimental in nature (than Studies 3 and 4) in terms of the degree of control of the learning experience. The participants were taught a pseudolanguage, which allows for maximal control of the variables of interest. Studies 3 and 4 were also experimental in nature but investigated naturalistic learners of L3 Swedish with L1 German and L2 English as the background languages. For Study 3, longitudinal data was collected stretching from beginner to CEFR C1 or higher, and for Study 4, cross-sectional data was collected.

The four studies in this thesis are outlined in Figure 40 below with respect to the type of learners, learning, acquired language, task type, and modifying factors.

	STUDY 1	STUDY 2	STUDY 3	STUDY 4
Learner	L1 English	L1 Swedish, L2 English	L1 German, L2 English	L1 German, L2 English
Learning	L2 Kontu v1 *artificial*	L3 Kontu v2 *artificial*	L3 Swedish *naturalistic, intensive*	L3 Swedish *naturalistic*
Forward CLI	L1 to L2 Accuracy (learning) Accuracy (outcomes) Processing (ToT)	L1/L2 to L3 Accuracy (learning) Accuracy (outcomes) Processing (ToT)	L1 to L2, L1 to L3, L2 to L3 Similarity evaluations Processing (ToT)	L1 to L2, L1 to L3, L2 to L3 Similarity evaluations Processing (ToT)
Reverse CLI	L2 to L1 Processing (RT)	L3 to L2, L3 to L1 Processing (RT)	L3 to L2, L3 to L1, L2 to L1 Similarity evaluations Processing (ToT)	L3 to L2, L3 to L1, L2 to L1 Similarity evaluations Processing (ToT)
Modifying factors	Progress, cognitive control, working memory, consolidation, retention	Progress, cognitive control, working memory, psychotypology	Progress, cognitive control, proficiency (L1, L2, L3), psychotypology	Location, psychotypology, proficiency (L1, L2, L3), age of onset (L2, L3), length of exposure (L1, L2, L3)

Figure 40. Overview of Studies 1-4 (learner, learning, direction, task type, and modifying factors)

In addition to the aforementioned modifying factors proficiency, aptitude, and psychotypology, CLI in the MML is subject to variation in aspects that are neither one of the outcome variables in the present thesis nor one of the modifying factors of interest. Every attempt has been made to control the effects of such factors. The Parasitic Model (Hall & Ecke, 2003) proposes thirteen factors that affect CLI in the MML. These can be divided into five broader areas, presented in Table 39 below. The empirical work in the present thesis covers all five areas.

Table 39

Controlled, measured, and manipulated aspects affecting CLI

AREA	FACTOR	Study 1	Study 2	Study 3	Study 4
Learner	Psychotypology		Measured	Measured	Measured
	Aptitude	Measured	Measured	Measured	
Learning	L2 status		Controlled	Controlled	Controlled
	Proficiency	Measured	Measured	Measured	Measured
	Acquisition order	Controlled	Controlled	Measured	Measured
Language	Typology	Manipulated	Manipulated	Controlled	Controlled
Event	Language mode			Controlled	Controlled
	Task	Measured	Measured	Measured	Measured
Word	Form	Controlled	Controlled	Controlled	Controlled
	Concept	Manipulated	Manipulated	Manipulated	Manipulated
	Concreteness	Controlled	Controlled	Controlled	Controlled
	Frequency	Controlled	Controlled	Controlled	Controlled
	Competitor frequency			Controlled	Controlled

There are three different ways in which the presented factors have been treated in the present thesis. Some of the factors have been kept constant in the sense that they have been *controlled*. Other factors have been *measured*, meaning that there has been a particular component in a particular study that has collected data on this factor. Finally, some of the factors have been *manipulated*, meaning that experimental items have been designed to vary in these aspects.

The present general discussion covers *directionality* (RQ1) as well as the effect of stage of acquisition (RQ3) on directionality in Section 7.2 below. The effects of the *modulating factors* (RQ2) as well as the interrelationship of modifying factors and stage of acquisition (RQ3) are presented in Section 7.3 below.

7.2 Direction of CLI

7.2.1 Forward and reverse CLI

The first overall research question related to the *direction* of CLI. We can divide the directions investigated in the present thesis into *forward* and *reverse*. With respect to forward CLI, an influence of the L1 on later acquired languages is expected. Such assumption is usually deemed uncontroversial for the lexicon (Jarvis & Pavlenko, 2008). On the other hand, research on reverse CLI, i.e., the effect of subsequent languages on the speaker's pre-existing languages have not been researched to the same extent. A third possibility is that CLI is *multidirectional* in nature (Sharwood-Smith, 1989). The present thesis has departed from the assumption that there is at least a theoretical possibility that all languages of a speaker influence each other. In L2 acquisition, there are two potential directions of CLI, where as in L3 acquisition, there are six potential directions of CLI. Table 40 below outlines the investigated directions of forward and reverse CLI in the four empirical studies in this thesis.

Table 40

Six tested directions of influence in the present thesis

		L1	L2	L3
Source	L1	–	1 3 4	2 3 4
	L2	1 3 4	–	2 3 4
	L3	2 3 4	2 3 4	–

Note. On the right, the source languages are presented. On the upper row, the target languages are presented. The numbers in the cells represents the study in this thesis (Study 1-4).

When it comes to the effects of translation ambiguity in forward and reverse CLI, four themes are discussed in the present subsection: CLI in learning, forward CLI, reverse CLI, and attrition. The learning of items with translation ambiguity was investigated in Studies 1 and 2. Both forward and reverse CLI were investigated in all four empirical studies in this thesis. Finally, attrition is a relevant theme of discussion with respect to the results in Studies 3 and 4.

Previous research has found that learning items with translation ambiguity is more difficult than learning items without translation ambiguity (Degani &

Tokowicz, 2010b; Eddington & Tokowicz, 2013). The results in Studies 1 and 2 align with these findings to some extent. That is, the participants were able to acquire both words with and without translation ambiguity with high accuracy so any effects, at least in the present data, relate to processing. In both Studies 1 and 2, the initial stage learners showed effects of earlier acquired languages on their acquisition of form-meaning mappings in the pseudolanguage. However, for the *ab initio* monolingual L2 learners, these effects were primarily visible in assessment, while for the *ab initio* bilingual L3 learners, these effects took place already during learning. Furthermore, only effects in processing were found: neither learner group showed any accuracy effects in assessment. These results are not in line with the findings of Degani, Tseng, and Tokowicz (2014), who found effects of translation ambiguity in accuracy but not in processing. If we take the findings from Degani et al. and Studies 1 and 2 together, we can, though, ascertain that CLI caused by translation ambiguity seems to present itself in either accuracy *or* processing.

Learners that have already acquired another language than their L1 have been found to be better at novel word learning than monolinguals (Kaushanskaya & Marian, 2009; Papagno & Vallar, 1995; Van Hell & Mahn, 1997). This has been hypothesized to be due to greater sensitivity to semantic information (Kaushanskaya & Marian, 2012). Comparing the results from Studies 1 and 2, this seems to be the case. The bilingual learners in Study 2 were first slower than their monolingual counterparts. The additional effort during learning in Study 2 led to better learning outcomes with respect to processing. The results align with the findings of Bogulski, Bice, and Kroll (2018) who found that English-Spanish bilinguals were slower at learning tasks than English monolinguals but that the learning outcomes for the bilinguals were better. In Study 2, two potential explanations were offered: 1) better *metalinguistic awareness* compared to the monolingual participants in Study 1 in line with Ringbom (2007), and 2) *desirable difficulty* in line with Bjork (1994). The results of the present dissertation are unable to differentiate between the two explanations, so this remains an avenue for further research.

Table 41 below summarizes the results of the four empirical studies with respect to forward CLI. Note that the results, as explained before, are different for monolingual learners in Study 1 and the bilingual learners in Study 2. The effects of the L1 on the L2 in second language acquisition and the L1 and the L2 on the L3 in third language acquisition are discussed below Table 41.

Table 41

Overview of forward CLI in Studies 1 through 4

	Study 1			Study 2			Study 3		Study 4	
	learn	acc	ass	learn	acc	ass	eval	time	eval	time
L1 to L2			item	-	-	-	ToT	global	ToT	?
L1 to L3	-	-	-	item			ToT	item	ToT	item
L2 to L3	-	-	-				ToT	item	ToT	item

Note. On the upper row, *learn* refers to ToTs during learning, *acc* to accuracy in assessment, *ass* to ToTs in assessment, *eval* to WSPT similarity evaluations, and *time* to ToTs in WSPT similarity evaluations. White cells with dashes represent untested directions of CLI and with question marks inconclusive data. Cells with shading in pink represent directions where no CLI was found, green directions where statistically significant effects were found, and beige directions where effects that were not statistically significant were found. Item and global refer to indications of item-level and global-level inhibition effects. ToT refers to cases where the overt evaluations were modulated by time spent on task.

For the effects of the L1 on the L2, there are three relevant models of development in the multilingual mental lexicon that make predictions with respect to CLI. The RHM predicts CLI from the L1 to the L2 immediately at the initial state. It does not however make predictions about translation ambiguity. The RHM-TA predicts differences in form-ambiguous and meaning-ambiguous translations. At the initial state, the predictions for the two types of ambiguity are the same, but at higher proficiency, more independence is predicted for the meaning-ambiguous (that is semantically unrelated) translations. The RHM-TA predicts that more automatized and more target-like processing develops in a linear fashion as the proficiency in the L2 increases. Finally, Jiang's (2000; 2002) L1 Lemma Mediation Hypothesis aligns mostly with the RHM (i.e., not the RHM-TA) in its predictions that L2 forms are mediated via the L1, albeit there is a hypothetical end-state of independence. The most important prediction that Jiang (2002) makes is that resolving translation ambiguity takes place using conscious rules that are not a part of the mental lexicon. This proposal entails that we should see a relationship between accuracy and processing in resolving translation ambiguity where one or the other can be observed. The prediction of all three models that there is L1 to

L2 CLI at the very initial state was corroborated by the results in Study 1. The results also aligned with the Parasitic Model when adjusted for L2. Furthermore, Jiang's hypothesis of conscious rules was corroborated by the results in Studies 1 and 3, and is discussed separately under Subsection 7.2.2 below.

For the effects of the L1 and the L2 on the L3, there is one relevant model of development in the MML: the Parasitic Model by Hall & Ecke (2003). Neither the L1 nor the L2 enjoys a privileged status as the primary supplier in the L3 in the PM. The choice between the two takes place at the item level based on salient properties of the particular word. In the present thesis, a conscious choice was made to not disambiguate between the effects of the L1 and the L2 on the L3 to ensure comparability in the reverse condition in Study 2. The observed results aligned with the PM and it was observed that the learners were able to detect differences, rehearse new patterns already during learning, and revise the representations without consolidation taking place in between. In the more naturalistic language learners in Studies 3 and 4, forward CLI was found both in processing and outcomes, with performance in outcomes being modulated by time on task (ToT). Longer ToTs co-occurred with more target-like performance in terms of outcomes. For Study 3 these effects were statistically significant whereas for Study 4 this was not the case. This is likely due to two reasons related to experimental design: 1) in Study 3, the participants acted as their own controls due to the longitudinal design, and 2) there was less between-subject variation in background variables in Study 3. Both aforementioned explanations have implications for statistical modeling. Before drawing any conclusions from comparing results from Studies 1 and 2 with data from Studies 3 and 4, it should be noted that the studies used different tasks. In the first two studies, effects in accuracy were limited to binary performance in a task with a correct and an incorrect response, whereas in Studies 3 and 4, the outcome variable was a task where change in similarity evaluations over time was used as the outcome variable. The latter allows for a more fine-grained estimation of change than the binary data in Studies 1 and 2, which potentially can lead to more sensitivity to conceptual CLI.

With respect to reverse CLI, Jarvis & Pavlenko (2008) suggest that the status of a L2/L3 language as a donor language in CLI is more related to factors such as proficiency, recency, and cross-linguistic similarity, than order of acquisition. In Studies 1 through 3, the order of acquisition in the data aligns with proficiency meaning that the L1 has the earliest AO, longest LoE, and highest proficiency, followed by the L2, and then the L3. In Study 4, there is between-participant variation in these. Study 4 was partially designed to address this question.

However, a methodological pitfall in Study 4 was that there was too much variation in the different subcomponents of proficiency and learning experience. No statistically significant effects of reverse CLI were found in Studies 1 and 2, so it is difficult to say what the effect of cross-linguistic similarity would be. In Studies 3 and 4, all three languages were very similar from the perspective of genealogical distance.

Some indications of CLI in all three potential directions of reverse CLI were found. An overview of the results of reverse CLI in Studies 1 through 4 is presented in Table 42 below.

Table 42

Overview of reverse CLI in Studies 1 through 4

	Study 1	Study 2	Study 3		Study 4	
	rt	rt	eval	time	eval	time
L3 to L2	-		ToT	global	ToT	global
L3 to L1	-			global		global
L2 to L1		-		?		

Note. See Table 39 above for explanation. The additional variable *rt* refers to response times in the masked priming task.

When it comes to reverse CLI, one aspect of individual variation in the participants in Studies 3 and 4 was *immersion*. Higby et al. (2020) hypothesized that immersion in the target language environment causes regulation at the global level (as opposed to item level) causing slower access in the previously acquired languages. This hypothesis is partially supported by the data from Study 3. Since the outcome task in Study 3 allowed for introspection, it is impossible to reliably differentiate between effects of regulation and introspection as both are significant in the data. In the case of forward CLI from L1 to L2, there were no significant effects of item type in ToTs. It should be remembered that the participants were proficient speakers of the L2, with long LoEs, already prior to the commencement of the data collection in Study 3. Thus, any effects of immersion could be postulated to apply for the immersed L3 only. Furthermore, Higby et al. (2020)

suggest that co-activation of translation equivalents causes a boost in the overall resting levels (i.e., not in a given moment) in both the target item (i.e., any word in a language) and its translation equivalents (in other languages). This could potentially explain the otherwise rather puzzling data about CLI from L3 to L2 in Study 3 (see Figure 32). In this condition, translation ambiguity facilitates processing of the L2 and the effect in the L2 increases in line with increasing proficiency in the L3. That is, for two items in the L2 that share a single translation equivalent in the L3, the participants' ToTs kept on reducing throughout the whole data collection period when compared to baseline items where each word had a separate translation equivalent in the L2 (and the L1). While this data could be explained by Higby and colleagues' hypothesis, the effect of L3 immersion on the L2 is a particularly unresearched area. Future research should aim to investigate the effects of L3 immersion on not only L1 attrition but also any effects that can be observed in the L2.

One aspect that is inherently connected to reverse CLI in L2/L3 immersion is L1 attrition. A temporary nature of effects in L1 attrition have been found using qualitative means (Opitz, 2013), quantitative methods (Linck, Kroll & Sunderman, 2009), and computational models (Meara, 2004, 2006). The data in Study 3 could potentially have been able to corroborate these findings. However, this was not the case. We can hypothesize that despite the participants having reached a high proficiency (CEFR C1 or higher) during the 10-month data collection period, the period was too short for any temporary "hit" on the L1 to have normalized. Further research should test intensive immersed language learners for periods longer than ten months.

If we, for a while, put aside the fact that the results in Studies 1 and 2 with respect to the effects observed in response times in the previously acquired languages after having acquired a new "language" were not significant, and take the observed effects at face value, we can hypothesize about what those effects would entail. In both Studies 1 and 2, there was an initial increase in the adjusted response times in the immediate post-test for those words in the pre-existing languages that had translation ambiguity between the pre-existing languages and the newly acquired pseudolanguage. This aligns with lexical competition (e.g., Green, 1998). In Study 1, which included a one-night consolidation period, and a subsequent post-test after that, we could observe that there was a reduction in the adjusted response times instead after consolidation. This aligns with facilitation (e.g., Costa, Miozzo & Caramazza, 1999). Previous research on early effects of reverse CLI point out that it may be that in addition to the proposed distinction between (response) accuracy and processing (time) that has been found in all four studies in the present thesis, it may be the case that to tap into the very initial effects, more sensitive research instruments should be used. Bice & Kroll (2015) found that (very) early effects of reverse CLI from the L2 Spanish to the L1 English in a lexical decision task were only observable in electrophysiological measures and not in accuracy or response times. Thus, further research with more sensitive research instruments, like ERPs, should be performed on third language acquisition as well testing reverse CLI in both the L2 and the L1 at the initial state.

It was pointed out that there is a shortage of models making predictions for three of the potential directions of CLI in third language acquisition. It is not unfounded to expect that this thesis would have put forth such a model, attempting to make predictions from the results of the six investigated directions of influence. This model would have tried to account for multidirectional CLI in the MML. However, there are of course people that speak more than three languages. The amount of potential directions increases exponentially with the addition of more languages. Perhaps the field does not need yet another "model" to be able to account for reverse CLI. Instead, research could focus on attempting to model multilingual lexical processing in a way that does not relate to specific language constellations in relation to order of acquisition. Focusing on the modifying factors (order of acquisition would, naturally, be one of them) it would be possible to make predictions that can be generalized to account for both within-participant, between-participant, within-language, and between-language variance. While the PM also has a component that models (forward) lexical development in third language acquisition, it is perhaps the *thirteen modifying factors* that the PM proposes that would be better suited as a starting point for modeling the "missing" directions of CLI. This thesis has looked at some of the potential modifying factors, attempting to disentangle their effect. One potential way forward, then, would be to not separately model, say, L5 to L2 influence, but future research of CLI in the MML could focus on the effect and interaction of the different modifying factors.

As a final remark on the observed effects of CLI in the experiments in the present thesis, we should maybe consider to which extent we can discuss languages as individual entities. The main outcome variable in this dissertation is cross-linguistic *influence* and the research designs, the analyses, and the discussion have focused on to what extent the languages affect each other. From a single-system hypothesis (see Section 2.1.2 as well as Kroll & Tokowicz, 2015 for an overview of the debate), though, we need to consider that the representations of the lexical items in the participants are not stable and, with the exception of Study 1, most likely not monolingual-like. The philosophical questions, then, are 1) whether we should primarily consider CLI as a metaphor for an analysis of the effects of languages in contact, and 2) if, and at which point, we should stop talking about CLI and consider that the representations have *converged* to form an intermediate language system (e.g., Alferink & Gullberg, 2013; Ameel, Malt, Storms & Van Assche, 2009; White, Malt & Storms, 2017) that differs from all of the speaker's languages' monolingual norms.

7.2.2 Introspection

The effects of CLI have been proposed to be more profound in situations where language learners are dependent on implicit knowledge (Jarvis, 2003; Jarvis & Pavlenko, 2008). On the other hand, a lion's share of research on CLI in the MML has focused on production errors (e.g., Bardel, Gudmundson & Lindqvist, 2012; Cenoz, 2001, 2003; De Angelis, 2005; De Angelis & Selinker 2001; Dewaele, 1998; Ecke, 2001; Hall & Ecke, 2003; Hammarberg, 2001, 2009; Jarvis, 1998; Lindqvist, 2009, 2010; Lindqvist & Falk, 2014; Neuser, 2017; Ringbom 1987, 2001; Singleton, 1987; Singleton & ó Laoire, 2006).

Jiang (2002) has hypothesized that second language learners often depend on *explicit* knowledge in resolving translation ambiguity. To give an example, Swedish does not lexicalize any difference between TO LEND and TO BORROW whereas such distinction is obligatory in English. Based on Jiang's predictions, a Swedish learner of English would have to consciously remember to adjust their use of the two English words in production. In this case CLI would take place at the level of representation, but upon successful application of the conscious rule, it might not represent itself in production. This type of conscious adjusting of one's language use has been referred to as *introspection* in this thesis.

In the results of the present thesis, indications of effects that are compatible with Jiang's hypothesis were found in all three directions of forward CLI in L3 acquisition: longer ToTs aligned with more target-like outcomes. Similar effects were found for L3 to L2 reverse CLI. Naturally, this does not entail that this is the *only* way language learners deal with translation ambiguity. Rather, we can hypothesize that this represents a secondary control mechanism, the first being inhibition. The methodological issue, here, is that both lexical competition and inhibition (in line with Elston-Güttler & Williams, 2008), and on other hand introspection (in line with Jiang, 2002) in the research instruments that have been used in the present study result in longer responses. Presumably, in the case of the latter, the observed increase in processing cost would be larger (with explicit rule application taking more time) than in the former case. In both cases we expect to see a relationship between response times and accuracy. In the latter case, time spent on applying explicit knowledge leads to more target-like outcomes, whereas in the former case co-activation of lexical and conceptual information in the other language could potentially result in less-target like outcomes. In any case, the results do highlight the importance of considering that learners, given favorable situational constraints, are able to adjust their language use despite CLI having taken place. Thus, focusing solely on production errors in outcomes may distort the picture with respect to what effects we observe and more importantly, fail to observe.

Finally, there is an interesting analogy to point out with respect to the suggestion that research on CLI may have focused too much on errors. De Bot (2004) argued that research on processing in the multilingual mental lexicon has focused too much on perception rather than production. This discrepancy results in there being vast amount of knowledge on perception of non-ambiguous words in multilinguals as well as a wealth of research in the outcomes of translation ambiguous words in production but a gap remains in combining the two.

7.3 Modulating factors

7.3.1 Proficiency

The first modulating factor representing *individual variation* in the subject pool was *proficiency*, a variable that inherently varies between language learners. Previous research has come to the conclusion that, by default, the amount of CLI decreases as the proficiency in the TL increases (e.g., Lindqvist, 2009; Navés et al., 2005; Neuser, 2017; Singleton, 1987; Williams & Hammarberg, 1998). However, even at near-native fluency, CLI does not seem to completely disappear (Abrahamsson & Hyltenstam, 2009; Birdsong, 2006; Lardiere, 2007). In the four studies in the present thesis, CLI was found in both less and more proficient learners. With respect to reduction of CLI as a result of increasing proficiency, the results indicate that the *type* (accuracy vs. processing) of CLI changes as a result of increased proficiency. That is, there is reduction in both, but even if effects in accuracy can no longer be observed, effects in processing are still present. This can be accounted for by cross-language activation.

With the exception of Study 4, the attempt has been to minimize between-subject variation in proficiency in both pre-existing languages as well as the language the learner is acquiring. In Studies 1 through 3, the participants were situated in a highly standardized language learning experience. In Studies 1 and 2, the learning period was very short, representing approximately one hour, and in Study 3 it was very long, representing 10 months. It is, however, not possible to completely eliminate between-subject variation in proficiency, and thus proficiency data was collected in Studies 1 through 3. It was, however, progress (in terms of time) that was used as the primary operationalization of proficiency in the first three studies. The results of generic effects of proficiency, not subject to individual variation, were discussed under forward CLI and reverse CLI above.

An analysis of the different components of proficiency representing individual variation in the learning experience and outcomes were, then, meant to be analyzed in Study 4 where, purposefully, data from participants with between-subject variation in the proficiency-related background factors were collected. Both subjective and objective measures were used. The subjective measures included self-reports of use, recency, age of onset (AO), length of exposure (LoE), manner of acquisition, as well as estimations of various skills (e.g., speaking, reading, writing, and listening) in a particular language. Objective measures were collected using a vocabulary task from the DIALANG as a proxy as it has been found to be a reliable approximation of proficiency (Alderson, 2005). For the languages of interest (i.e., the ones that act as either as a source or a target language), objective measures were collected in all four studies.

The results in Study 4 align with the findings from Study 3 but it was not possible to disentangle the effects of the different subcomponents of proficiency with the number of participants (*n* = 29) the study had. Data collection for Study 4 had to be halted in spring 2020 due to health concerns. However, even if there had been, say, ten more participants, this would not have improved the predictive power. While Study 4 was not able to provide results for those questions that related to the different subcomponents of proficiency, it was by all means not done in vain. The results of Study 4 compared to those in Study 3 provided insights into the effects (and non-effects) of immersion and intensive language learning experience, and the effects of psychotypology (presented in 7.3.3 below).

Study 4 has also provided us with probably the most intriguing finding about proficiency. The best predictor (*r* = 0.59) of a high objective vocabulary score in the L1 German was a high objective vocabulary score in the L3 Swedish. There are multiple possible explanations for this. The most simple one is that some learners are good at vocabulary tests. Another possibility is that some learners work more actively to improve their vocabulary in all of their languages. The third explanation was previously offered for unexpected results in Study 3: Higby et al. (2020) have proposed that in addition to a *direct frequency effect* caused by the activation of particular vocabulary items in a particular language, the translation equivalents of those items see an *indirect frequency effect*. The latter would likely be most profound when the language being used is a later acquired, less proficient language, and the other language is the mother tongue. This would be the case due to lexical access via the L1 (as proposed by the RHM, The L1 Lemma Mediation Hypothesis, and the PM). Let us hypothesize that we have two speakers (S1, S2) of a L3, both immersed in the L3 environment. Both speakers use the same amount of the L3, the L2, and the L1 with the L3 being the most commonly used language. The length of residence for both speakers is the same. The difference between these two speakers is that S1 uses lexically diverse language in the L3 and S2 uses *less* lexically diverse language in the L3. We could then hypothesize that due to the indirect frequency effect, a larger "proportion" of the low frequency items in the L1 would be accessible despite them not being used as a result of frequent co-activation from the L3 in S2. This, again, highlights the importance of disentangling the effects of cross-language activation and attrition.

7.3.2 Aptitude

The second modulating factor in this thesis was *aptitude* which is a set of predispositions that predict a language learner's performance under a set of given conditions. It has been suggested to have a larger impact on late language learners (Ortega, 2013). *Cognitive control* (e.g., Green, 1998) and *working memory* (e.g., Papagno & Vallar, 1992; Service & Kohonen, 1995) have been proposed as predictors of success in vocabulary learning and processing (Ellis & Sinclair, 1996). Sometimes teasing apart the effects of proficiency and aptitude can be problematic if higher aptitude leads to better learning results, increased proficiency to more available resources, and more available resources to additional learning.

In the present thesis, aptitude has been operationalized as cognitive control (CC) and working memory (WM), as well as self-evaluated measures of language learning aptitude, attitude, and the ability for cross-cultural communication. Self-evaluations were collected in all four studies. In Studies 1 and 2, both WM and CC were measured. In Study 3, CC data was collected and in Study 4 there is only self-reported data. With the exception of Study 4, the participants in this dissertation were learning the L2/L3 (i.e., the language in which learning is being observed) as adults. In Study 4, AO is used as a modulating factor.

In line with the hypothesis that success in late learners is more affected by aptitude (Ortega, 2013), we would expect the participants in Study 1 to be most affected. For reference, Table 43 below presents correlations of all aptitude measures in Study 1. Additional information is provided in Appendix E. Table 43 shows that the correlations within self-reported measures are higher than those between self-reported measures and more objective measures. This highlights the importance of gathering objective data.

Table 44 below summarizes the results of WM and CC as modulating factors in Studies 1 through 3. No objective aptitude measures were collected in Study 4. Effects of WM were only found in Study 1 which aligns with the hypothesis of late learners being more affected. No effects of CC were found in Study 1. The participants in Study 2, while at the initial state in the L3, had already successfully acquired two languages and CC was a modulating factor. This likely relates to the participants in Study 2 having a) two languages to regulate, and b) having acquired the ability to regulate languages, neither of which can be postulated to be the case for the participants in Study 1.

Table 43

Correlation matrix for aptitude measures

		APT	ATT	COM	CCFLA	CCAC	W2RT	W2AC	W3RT	W3AC
APT	*r*	–								
	p	–								
ATT	*r*	0.738	–							
	p	< .001	–							
COM	*r*	0.526	0.676	–						
	p	0.003	< .001	–						
CCFLA	*r*	0.255	0.236	0.276	–					
	p	0.174	0.210	0.140	–					
CCAC	*r*	0.389	0.519	0.416	0.341	–				
	p	0.034	0.003	0.022	0.066	–				
W2RT	*r*	-0.115	-0.028	0.205	-0.292	-0.165	–			
	p	0.547	0.884	0.278	0.117	0.385	–			
W2AC	*r*	0.150	0.004	-0.131	-0.059	0.148	0.123	–		
	p	0.428	0.981	0.491	0.757	0.435	0.517	–		
W3RT	*r*	-0.111	-0.178	0.057	-0.337	-0.357	0.875	0.331	–	
	p	0.559	0.347	0.763	0.069	0.053	< .001	0.074	–	
W3AC	*r*	0.406	0.414	0.347	0.361	0.238	0.048	0.371	0.057	–
	p	0.026	0.023	0.060	0.050	0.205	0.800	0.044	0.764	–

Note. Aptitude (APT), attitude (ATT), and intercultural communicative ability (COM) are self-rated. Cognitive control is represented by two measures: flanker effect (CCFLA) and accuracy in the flanker task (CCAC). W2 and W3 refer to the 2-back and 3-back tasks variant of the *n*-back task.

Even in the highly multilingual participants in Study 3, CC was only a relevant predictor of reverse CLI. One possibility is that the effect of aptitude varies not only between L1 and L2 acquisition but also between L2 and L3 acquisition. L2 learners likely do not have the same amount of metalinguistic knowledge about their L1 as L3 learners have about their L2. Furthermore, it could be hypothesized that L3 learners can put *analytic* abilities to better use. The role of language analytic ability was recently highlighted by a longitudinal study of the effects of different subcomponents of aptitude in child language learners (Roehr-Brackin & Tellier, 2019). It could be hypothesized, that adult learners would be better able to put analytic abilities to use. Such data was, however, not collected for the four studies in the present dissertation.

Table 44

Overview of aptitude in Studies 1 through 3

	Study 1		Study 2		Study 3	
	WM	CC	WM	CC	WM	CC
L1 to L2	A		-	-	-	
L1 to L3	-	-		O	-	
L2 to L3	-	-			-	
L3 to L2	-	-			-	
L3 to L1	-	-		RT	-	ToT
L2 to L1			-		-	

Note. White cells with dashes represent directions in which the effect of aptitude has not been tested. A relates to ToTs in assessment, O to accuracy in outcomes, RT to response times in the priming task, and ToT to time on task in the WSPT similarity assessments.

Finally, in previous research, Link, Hoshino, and Kroll (2008) found that better WM and CC relate to reduced cross-language activation in bilinguals. Their results suggested that WM is more important in comprehension whereas CC is more important in production. The outcome variables in the studies in the present dissertation are neither strictly related to comprehension nor production albeit more geared towards comprehension. All the outcome tasks include comprehension of linguistic stimuli (one or more words) as well as require the learner to do something with that information (decide if the stimulus is a real word, choose a correct meaning from a set of options, or evaluate similarity in meaning).

7.3.3 Psychotypology

The third, and last, modulating factor that was of interest in the present thesis was *psychotypology*, which refers to *perceived* language distance (Kellerman, 1983). The Parasitic Model (Hall & Ecke, 2003) makes a difference between psychotypology as a learner factor and *typological distance* as a language factor. In previous research, psychotypology has been treated both as a *static* factor (e.g., Flynn, Foley & Vinnitskaya, 2004; Rothman, 2015) and a *fluid* factor (Neuser, 2017; Rast, 2008;

Sayehli, 2013; Suhonen, 2015; Xia, 2017). In the geneologically-based approach, similarity can be predicted based on relevant properties of the languages in question. When psychotypology is treated as a fluid factor, it has to be measured in each individual. Furthermore, it has been proposed that there is a difference between *assumed* and *perceived* psychotypology (Kaivapalu, 2004; Kaivapalu & Muikku-Werner, 2010). Assumed psychotypology refers to the learner's preconceptions about similarity between two languages, while perceived psychotypology relates to the learner's evaluations of similarity based on exposure. Finally, if psychotypology is measured – as opposed to predicted based on relevant properties – it can be measured both *explicitly* and *implicitly*. In the case of the former, the instrument should allow for the use of conscious knowledge, like metalinguistic awareness, whereas in the latter case, the instrument should tap into the learner's unconscious performance.

In the present thesis, psychotypology was treated as a fluid factor that is subject to within-subject and between-subject variation. The benefit of treating psychotypology as a construct with within-subject and between-subject variation is that it includes the *possibility* that psychotypology is static, meaning that there is no between-subject or within-subject variation over time that cannot be attributed to measurement error. The caveat is that one has to rely on the reliability of the measuring instruments. Furthermore, psychotypology was measured both explicitly and implicitly. For the explicit measurement of psychotypology, the participants were explicitly asked about their perceptions of similarities (and differences) between languages. For its implicit measurement, the participants' willingness to accept items derived from one of the potential source languages was tested instead. Furthermore, in Study 2, data on both assumed and perceived similarity was collected. In the present thesis, two instruments were used to measure psychotypology. The Explicit Cross-Linguistic Similarity Assessment task (E-CLSA) was used in Studies 2 through 4. In Study 2, it was used to probe assumed similarity and in Studies 3 and 4, it was used to probe perceived similarity. Furthermore, the Implicit Cross-Linguistic Similarity Assessment task (I-CLSA) was used in Study 2 to probe perceived similarity.

Little correlation was found between the measures in the two types of tasks in Study 2 where they were both used. This aligns with the hypothesis of Kaivapalu (2004) that there is a difference between psychotypology that is assumed (i.e., that a learner has before learning a language) and that is perceived (i.e., that the learner has after *x* amount of exposure to the language). In Study 2, the amount of exposure in the pseudolanguage was approximately one hour, which can, potentially, have implications for how reliable a measure of perceived similarity is, depending on how much input is required to establish perceived similarity.

Further research should attempt to disentangle the extent to which assumed and perceived similarity vary, as well as whether perceived similarity fluctuates after exposure to the L2/L3. A methodological shortcoming in Study 2 with respect to this particular question is that different instruments were used to probe the two types of psychotypology, as the alignment of the learner's conscious and unconscious behavior was also of interest. A differentiation between assumed and perceived, and on the other hand conscious and unconscious, could have been achieved by administering both task types prior to and after exposure to the L3 Kontu v2. The reason for why this was not done is two-fold. First, it would have been difficult to administer the I-CLSA prior to learning without it affecting the expectations with respect to the language that was about to be learned. Second, all stimuli up until the administering of I-CLSA after learning were carefully adjusted so that no formal similarities existed between the pseudolanguage and the pre-existing languages, because such similarities could have caused form-driven learning effects on particular items (Jarvis, Raines, Schaefer & Sormaz, 2019).

Furthermore, a discrepancy was observed with respect to the effect of psychotypology in Study 2 in that the results for forward and reverse CLI were different. Neither E-CLSA, which was administered *prior to* learning, nor I-CLSA, which was administered *after* learning, predicted accuracy or processing during learning. There are three potential explanations for this, which are outlined below.

The first explanation is that for *x* portion of the time spent on learning, the learners had not had enough input to have established *perceived* similarity. Rothman (2015) refers to a *transitory stage* during which a learner goes through a comparative process. If the length of this process aligned with the time it took the participants to learn the pseudolanguage, there would be no perceived psychotypology-driven preferences. In this case, the lack of explanatory power would also entail that *assumed* similarity does not predict learner behavior since it had already been established.

The second explanation has to do with the genealogical relationship between Finnish, English, and Swedish as previous research has found that CLI tends to take place from the typologically closer language (e.g., Cenoz, 2001, 2003; De Angelis, 2005; De Angelis & Selinker, 2001; Leung, 2005; Lindqvist, 2009; Lindqvist & Bardel, 2013; Odlin & Jarvis, 2004; Ringbom, 1987; Singleton & Ó Laoire, 2006). With English and Swedish as pre-existing languages, it could be the case that the learners simply found the Finnish-based Kontu v2 too distinct. A related potential explanation is that while Swedish and Finnish do share some vocabulary, any surface overlap between the Kontu v2 forms and the pre-existing languages was avoided. The reason for this methodological choice was once again to avoid form-based effects in learning since one of the main objectives of the study was to compare the learning of words depending on manipulations of their meaning.

The third explanation is that psychotypology does not predict performance in the learning of a pseudolanguage. That is, the learners did not treat Kontu v2 as a natural language. In this case we would not expect any reverse effects (i.e., influence from the later acquired languages into the previously acquired languages) either. However, perceived similarity in the unconscious task modulated reverse CLI with higher perceived similarity between English and the pseudolanguage resulting in increased overall reverse CLI. In this case, it seems that the inhibition effort was modulated by perceived similarity, which seems to suggest that at least *after* the learning of Kontu, the participants had acquired some form of perceived representation of language similarity rebuffing the third explanation.

All in all, it seems that the first explanation, i.e., that of a transitory stage, is most plausible. This would entail two things. First, *assumed* similarity does not predict (unconscious) performance in learning, at least after some input. It may be the case, however, that assumed similarity could, for example predict *inferencing* as a conscious process. Second, during any postulated transitory stage, no reliable predictions can be made from a *perceived* similarity measurement that was collected after the postulated transitory stage. Since the account of *assumed* and *perceived* similarity refers to estimations of similarity *before* learning and *after* input, the account is not necessarily incompatible with the Typological Primacy Model which, though, makes no predictions about outcomes in vocabulary (Rothman, 2015; Rothman et al., 2019).

In Study 3, conscious evaluations of psychotypology predicted behavior in the L1 German. Higher perceived similarity between either SL (L2 English or L3 Swedish) and the L1 German led to shorter ToTs in the WSPT task. Furthermore, CC and psychotypology had a statistically significant interaction in that the combination of better CC and higher perceived similarity also led to shorter ToTs. This, presumably, means that the participants' behavior that is modified by psychotypology is also dependent on their available cognitive resources. This entails that a higher perceived similarity would lead to less applied inhibition, which is expected based on previous findings indicating that more similar languages are more likely to activate each other than less similar languages (Ringbom, 2007).

An unexpected effect of psychotypology was found in the L2 English in both Study 3 and 4. Higher perceived similarity of either SL (L1 German or L3 Swedish) and the TL L3 English led to longer ToTs in the WSPT task. These results likely represent some form of hypercorrection or *avoidance* (Schachter, 1974), but since the experiments were not specifically designed to test avoidance it is difficult to say whether there can be other explanations. What speaks in favor of the avoidance strategy is that the participants' L2 English, in neither Studies 3 or 4, was a language that the participants were actively learning. However, English was a language that the participants were very proficient in. It was also the case that the three languages in question (German, English, and Swedish) are typologically very similar to each other. With respect to cross-linguistic activation due to form, the learners' have every reason to be wary of influence of the L1 German and the L3 Swedish in the L2 English.

To summarize, given the combination of the low correlation between the two psychotypology measures, as well as the fact that both predicted the learners' behavior, we must conclude that further research is needed in disentangling the components and effects of psychotypology.

7.4 Suggestions for further research

The present thesis represents a unique constellation of studies on CLI in late foreign language learners in that it presents data 1) covering a broad range of proficiencies, 2) in all six potential directions in third language acquisition, and 3) on both accuracy and processing. It does, however, have some limitations, as outlined in the following.

The experiments in this thesis have looked at the lexicon in isolation. Hopefully, the results can provide a starting point for further research on the developmental aspects of the multilingual mental lexicon. However, to align the developmental aspects of lexical knowledge with other areas of linguistic knowledge, longitudinal research is needed on both forward and reverse CLI in the lexicon in relation to other areas of linguistic knowledge in the same population and the same study.

The time periods in Studies 1 through 3 were too short to make any definitive suggestions concerning which effects are temporary by nature and which not. Further research should add additional data points to make the studies proportionally longer, i.e., days to weeks if replicating Studies 1 and 2, and months to years if replicating Study 3. With respect to modifying factors, studies with much larger sample sizes are required to properly disentangle between the different background variables.

Finally, and perhaps most importantly, future research should aim to differentiate between the effects of attrition and CLI. I envision interesting future research disentangling the effects of attrition that are caused by lack of activation, representational changes caused by activation of related information, temporary lapses in outcomes caused by availability (or lack thereof) of cognitive resources, (temporary) processing issues caused by intensive acquisition with regulation skills lagging behind, and unavoidable effects of CLI at the level of representation due to spreading activation. All of the above have the capacity to lead to the same overt outcomes.

8 Conclusion

This dissertation has investigated cross-linguistic influence (CLI) in the developing multilingual mental lexicon (MML) in adults. CLI in the MML was found to be, to some extent, multidirectional. Forward CLI was found across the board, modulated by introspection. Statistically significant reverse CLI – that cannot be accounted for by attrition – was found from the L3 to the L2 in the intensive language learners in Study 3. Furthermore, some indications (lacking statistical significance) of reverse CLI were found in all potential directions as well as in the early and late stages of acquisition. Further research, potentially using more sensitive instruments, such as ERPs, would be required to further investigate them.

CLI was found to be modulated by the three hypothesized factors proficiency, aptitude, and psychotypology, and these were interdependent with both the direction of CLI as well as the stage of acquisition. However, each factor had their individual issues. Analyzing the effects of the various components of proficiency would require more statistical power. Also, an investigation of longitudinal type, like the one in Study 3, should measure the different modifying factors at each data collection session repeatedly. Finally, little correlation was found between conscious and unconscious estimates of psychotypology, suggesting that future research should pay close attention to the operationalization of psychotypology. Finally, comparing data from Studies 3 and 4, it is evident that using participants as their own controls is preferred over cross-sectional designs.

The findings, overall, highlight the importance of considering the lack of conceptual non-equivalence in modeling multilingual lexical processing as well as the importance of separating the effects of attrition from the effects of reverse CLI.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*(2), 249–306.

Abutalebi, J., Cappa, S. F., & Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In J. Kroll, & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 497–515). Oxford: Oxford University Press.

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*, 242–275.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* London: Continuum.

Alferink, I., & Gullberg, M. (2013). French–Dutch bilinguals do not maintain obligatory semantic distinctions: Evidence from placement verbs. *Bilingualism, 17*, 22–37.

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language, 60*, 270–290.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*, 451–474.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior*, *22*(3), 261–295.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063–1087.

Anderson, R. C., & Nagy, W. E. (1989). *Word meanings.* Champaign: University of Illinois at Urbana-Champaign.

Ausubel, D. P. (1964). Adults versus children in second-language learning: Psychological considerations. *The Modern Language Journal, 48*(7), 420–424.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Bardel, C. (2015). Lexical cross-linguistic influence in third language development. In H. Peukert (Ed.), *Transfer effects in multilingual language development* (pp. 111–128). Amsterdam: John Benjamins.

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

Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 61–78). Amsterdam: John Benjamins.

Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production: Vocabulary acquisition and use in L2 French and Italian. *Studies in Second Language Acquisition, 34*(2), 269–290.

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychological 142,* 402–409.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3–11.

Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language acquisition. *Neuroreport, 16*, 966–971.

Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition, 42*, 23–60. Reprinted in W.J.M. Levelt (Ed.), *Lexical access in speech production*. Cambridge, MA: Blackwell.

Birdsong, D. (2006). Age and second language acquisition and processing. A selective overview. *Language Learning, 56*(1), 9–49.

Bjork, R. A. (1994). Institutional impediments to effective training. In D. Druckman, & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 295–306). Washington, DC: National Academy Press.

Bogulski, C. A., Bice, K., & Kroll, J. F. (2018). Bilingualism as a desirable difficulty: Advantages in word learning depend on regulation of the dominant language. *Bilingualism: Language and Cognition*. Advance online publication.

Boratynska-Sumara, J. (2014). Lexical transfer research in third language acquisition (TLA) – an overview. *Studia Linguistica Universitatis Iagellonicae Cracoviensis, 131,* 137–148.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston, MA: Houghton Mifflin.

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: Multilingual Matters.

Cenoz, J. (2003). The role of typology in the organization of the multilingual lexicon. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *The multilingual lexicon* (pp. 103–116). Dordrecht: Kluwer.

Cenoz, J. (2008). *Looking beyond second language acquisition. Studies in tri- and multilingualism.* Tübingen: Stauffenberg Verlag.

Cenoz, J., Hufeisen, B., & Jessner, U. (2001). *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives.* Clevedon: Multilingual Matters.

Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41*(3), 365–397.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283–1296.

Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2017). Do bilinguals automatically activate their native language when they are not using it? *Cognitive Science, 41*(6), 1629–1644.

Davies, M. (2010). The corpus of contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing, 25*(4), 447–464.

De Angelis, G., & Selinker, L. (2001). Interlanguage transfer and competing linguistic systems in the multilingual mind. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 42–58). Clevedon: Multilingual Matters.

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

De Angelis, G. (2005b). Multilingualism and non-native lexical transfer: An identification problem. *International Journal of Multilingualism, 2,* 1–25.

De Angelis, G. (2007). *Third or additional language acquisition*. Clevedon: Multilingual Matters.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, *13*(1), 1–24.

De Bot, K. (2004). The multilingual lexicon: Modeling selection and control. *International Journal of Multilingualism, 1,* 17–32.

De Groot, A. M. B. (2011). *Language and cognition in bilinguals and multilinguals: An introduction.* New York: Psychology Press.

Degani, T., & Tokowicz, N. (2010a). Semantic ambiguity within and across languages: An integrative review. *The Quarterly Journal of Experimental Psychology, 63*(7), 1266–1303.

Degani, T., & Tokowicz, N. (2010b). Ambiguous words are harder to learn. *Bilingualism: Language and Cognition, 13,* 299–314.

Degani, T., Tseng, A., & Tokowicz, N. (2014). Together or apart? Learning of ambiguous words. *Bilingualism: Language and Cognition, 17*, 749–765.

Dentler, S. (2000). Deutch und Englishsch – dasgibt immer Krieg!. In S. Dentler, B. Hufeisen, & B. Lindemann (Eds.), *Tertiär und Drittsprachen. Projekte und empirische Untersuchungen* (pp. 42–58). Tübingen: Stauffenberg Verlag.

Dewaele, J. M. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics 19*(4), 471–490.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. Kroll, & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–199). Oxford: Oxford University Press.

Domene Moreno, C. (2019). *Beyond transfer? The acquisition of an L3 phonology by Turkish-German bilinguals*. Unpublished doctoral dissertation. Julius-Maximilians-Universität Würzburg, Germany.

Ecke, P. (2001). Lexical retrieval in third language. Evidence from errors and tip-of-the-tongue states. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 90–114). Clevedon: Multilingual Matters.

Ecke, P. (2004). Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism, 8,* 321–354.

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: Language and Cognition*, *18*, 145–162.

Eddington, C., & Tokowicz, N. (2013). Examining English-German translation ambiguity using primed translation recognition. *Bilingualism: Language and Cognition, 16*, 442–457.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413.

Ellis, N. C., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology, 49*(1), 234–250.

Elston-Güttler, K. (2000). *An enquiry into cross-language differences in lexical-conceptual relationships and their effect on L2 lexical processing.* Unpublished doctoral dissertation. Cambridge University, United Kingdom.

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005). Who's in control?: Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience, 17*, 1593–1610.

Elston-Güttler, K. E., & Williams, J. N. (2008). First language polysemy affects second language meaning interpretation: Evidence for activation of first language concepts during second language reading. *Second Language Research*, *24*(2), 167–187.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge: Cambridge University Press.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149.

Falk, Y. (2010). *Gingerly studied transfer phenomena in L3 Germanic syntax. The role of the second language in third language acquisition.* Utrecht: LOT Publications.

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., & Bardel, C. (2010). The study of the role of the background languages in third language acquisition. The state of the art. *International Review of Applied Linguistics in Language Teaching. IRAL 48*(2-3), 185–220

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. *International Journal of Multilingualism, 1,* 3–17.

Fodor, J. D. (1977). *Semantics theories of meaning in generative grammar*. New York: Crowell.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680–698.

Fuster, C., & Neuser, H. (2020). Exploring intentionality in lexical transfer. *International Journal of Multilingualism,* 17(4), 516–534.

Gabrys-Barker, D. (2005). *Aspects of multilingual storage, processing, and retrieval.* Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Geeraerts, D., Bakema, P., & Grondelaers, S. (1994). *The structure of lexical variation: meaning, naming, and context.* Berlin: Mouton de Gruyter.

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek, & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). Oxford: Oxford University Press.

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

Green, D. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain And Language*, *27*(2), 210–223.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81.

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

Green, D., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience, 29*(4), 499–511.

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Press.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning, 60*(S2), 5–24.

Gyllstad, H., & Suhonen, L.-V. (2017). Is attrition a type of learning. *Linguistic Approaches to Bilingualism, 7*(6), 700–703.

Haastrup, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tubingen: Gunter Narr.

Haghverdi, H. R., Tabrizi, H. H., & Eghbalitabar, M. M (2012). The effectiveness of psychotypology-reduced L2 teaching on three linguistically different groups of Iranian undergraduate EFL learners' reading comprehension skill. *World Journal of English Language, 2*(1), 44–56.

Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.) *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 21–41). Clevedon: Multilingual Matters.

Hammarberg, B. (2009). *Processes in third language acquisition.* Edinburgh: Edinburgh University Press.

Hammarberg, B. (2018). L3, the tertiary language. In A. Bonnet, & P. Siemund (Eds.), *Foreign language education in multilingual classrooms* (pp. 127–150). Amsterdam: John Benjamins.

Hall, C. J. (2002). The automatic cognate form assumption: evidence for the parasitic model of vocabulary development. *International Review of Applied Linguistics, 40,* 69–87.

Hall, C., & Ecke, P. (2003). Parasitism as a default mechanism in L3 vocabulary acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *The multilingual lexicon.* Dortrecht: Kluwer Academic Publishers.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*(1), 25–38.

Higby, E., Donnelly, S., Yoon, J., & Obler, L. K. (2020). The effect of second-language vocabulary on word retrieval in the native language. *Bilingualism: Language and Cognition, 23*(4), 812-824.

Jarema, G., & Libben, G. (2007). Matters of definition and core perspectives. In G. Jarema, & G. Libben (Eds.), *The mental lexicon 1*(2) (pp. 1-6). Amsterdam: Elsevier.

Jarvis, S. (1998). *Conceptual transfer in the interlingual lexicon*. Bloomington: Indiana University Linguistics Club Publication.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning, 50*(2), 245–309.

Jarvis, S. (2003). Probing the effects of the L2 on the L1: A case study. In V. Cook (Ed.), *Effects of the second language on the first* (pp. 81–102). Clevedon: Multilingual matters.

Jarvis, S., & Pavlenko, A. (2008). *Cross-linguistic influence in language and cognition.* New York: Routledge.

Jarvis, S., Raines, T., Schaefer, P., & Sormaz, O. (2019). Effects of crosslinguistic similarity, complexity, and depth of processing on vocabulary recall. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 273–285). New York: Routledge.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47–77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*(4), 617–637.

Kachru, B. (1997). World Englishes and English-using communities. *Annual Review of Applied Linguistics, 17*, 66–87.

Kaivapalu, A. (2004). Kui sarnane on sarnane? Eesti ja soome mitmuse-vormide psühholingvistilisest reaalsusest. In H. Sulkala, & H. Laanekask (Eds.), *VIRSU II. Suomi ja viro kohdekielinä* (pp. 62-71). Oulu: Oulun Yliopisto.

Kaivapalu, A., & Muikku-Werner, P. (2010). Reseptiivinen monikielisyys: Miten suomenkielinen oppija ymmärtää viroa äidinkielensä pohjalta? *Lähivõrdlusi/Lähivertailuja, 20,* 68–97.

Kato, T. (2014). Development of the sleep quality questionnaire in healthy adults. *Journal of Health Psychology, 19*(8), 977–986.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*, 705–710.

Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin & Review, 19*, 935–941.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627–633.

Kellerman, E. (1983). Now you see it, now you don't. In S. Gass, & L. Selinker (Eds.), *Language transfer in language learning* (pp. 112–134). Rowley, MA: Newbury House.

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

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352–358.

Krashen, S. (1983). Newmark's "Ignorance hypothesis" and current second language acquisition theory. In S. Gass, & L. Selinker (Eds.), *Language transfer in language learning* (pp. 135–153). Rowley, MA: Newbury House.

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

Kroll, J., Bogulski, C., & McCain, R. (2012). Psycholinguistic perspectives on second language learning and bilingualism: The course and consequence of cross-language competition. *Linguistic Approaches to Bilingualism, 2*(1), 1–24.

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, Mind, and Brain. *Annual review of linguistics*, *1*, 377–394.

Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll, & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531–553). Oxford: Oxford University Press.

Laalo, K. (1990). Ruuvi, rossi, krossi ja skruuvi. Lainasanat ja suomen muuttuva äännejärjestelmä. *Kielikello 4/1990*. Helsinki: Institute for the Languages of Finland.

Langacker, R. (1987). Nouns and verbs. *Language, 63*(1), 53–94.

Lardiere, D. (2007). *Ultimate attainment in second language acquisition: A case study.* Nahwaw, NJ: Erlbaum.

Laxen, J., & Lavaur, J-M. (2010). The role of semantics in translation recognition: Effects of number of translations, dominance of translations and semantic relatedness of multiple translations. *Bilingualism: Language and Cognition, 13,* 157–183.

Lee, M-W., & Williams, J. N. (2001). Lexical access in spoken word production in bilinguals: Evidence from the semantic competitor priming paradigm. *Bilingualism: Language and Cognition, 4*(3), 233–248.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Leung, Y. (2005). L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese–English bilinguals. *Bilingualism: Language and Cognition, 8*(1), 39–61.

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

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science, 18*(1), 29–34.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673–680.

Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *Mental Lexicon, 3,* 349–374.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507–1515.

Lindqvist, C. (2009). The use of the L1 and the L2 in French L3. Examining cross-linguistic lexemes in multilingual learners' oral production. *International Journal of Multilingualism, 6*(3), 281–297.

Lindqvist, C. (2010). Inter- and intralingual lexical influences in advances learners' French L3 oral production. *International Review of Applied Linguistics in Language Teaching, 48*(2–3), 131–157.

Lindqvist, C. (2015). Do learners transfer from the language they perceive as most closely related to the L3? The role of psychotypology for lexical and grammatical crosslinguistic influence in French L3. In G. De Angelis, U. Jessner, & M. Kresić (Eds.), *Crosslinguistic influence and crosslinguistic interaction in multilingual language learning* (pp. 231–252). London: Bloomsbury Academic.

Lindqvist, C., & Bardel, C. (2013). Exploring the impact of the proficiency and typology factors: Two cases of multilingual learners' L3 learning. In M. Pawlak, & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism. Studies in honor of David Singleton* (pp. 253-266). Cham: Springer.

Lindqvist, C., & Falk, Y. (2014). When Germans begin to learn Swedish: Which is the transfer source for function words, content words and syntax? In L. Roberts, I. Vedder, & J. H. Hulstijn (Eds.), *EUROSLA Yearbook* (Vol. 14, pp. 225–239). Amsterdam: John Benjamins.

Lloyd-Smith, A. (2020). *Heritage bilingualism and the acquisition of English as a third language.* Unpublished PhD Thesis. University of Konstanz, Germany.

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

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

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*, 97–115.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. Part 1: An account of basic findings. *Psychological Review, 88*, 375–407.

Meara, P. M. (2004). Modelling vocabulary loss. *Applied Linguistics, 25*(2), 137–155.

Meara, P. M. (2006). Emergent properties of multilingual lexicons. *Applied Linguistics, 27*(4), 620–644.

Meisel, J. (1983). Transfer as a second language strategy. *Language and Communication 3,* 11–46.

Mencken, H. L. (1937). *The American language: An inquiry into the development of English in the United States.* New York: Knopf.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language, 67*(1), 224–237.

Murphy, G. (2002). *The big book of concepts*. Cambridge, MA: The MIT Press.

Naves, T., Miralpeix, I., & Luz Celaya, M. (2005). Who transfers more...and what? Cross-linguistic influence in relation to school grade and language dominance in EFL. *International Journal of Multilingualism, 2*(2), 113–134.

Neuser, H. (2017). *Source language of lexical transfer in multilingual learners.* Unpublished doctoral dissertation. Stockholm University, Sweden.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. E. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–14). New York: Plenum Press.

Odlin, T. (1989). *Language transfer. Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, T. (2003). Cross-linguistic influence. In C. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 436–486). Malden, MA: Blackwell.

Odlin, T., & Jarvis, S. (2004). Same source, different outcomes: A study of Swedish influence on the acquisition of English in Finland. *International Journal of Multilingualism*, *1*(2), 123–140.

Odlin, T. (2012). Cross-linguistic influence in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Oxford: Wiley-Blackwell.

Opitz, C. (2013). A dynamic perspective on late bilinguals' linguistic development in an L2 environment. *International Journal of Bilingualism, 17*(6), 701–715.

Ortega, M., (2008). Cross-linguistic influence in multilingual language acquisition: The role of L1 and non-native languages in English and Catalan oral production. *Íkala*. *13*(1), 121–142.

Ortega, L. (2009). Sequences and processes in language learning. In M. H. Long, & C. Doughty (Eds.), *The handbook of language teaching* (pp. 81-105). Chichester: Wiley-Blackwell.

Ortega, L. (2013). *Understanding second language acquisition.* London: Routledge.

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *Quarterly Journal of Experimental Psychology, 44*(1), 47–67.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, *48A,* 98–107.

Paradis, C. (2012). Lexical semantics. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.

Paradis, C. (2015). Meanings of words: Theory and application. In U. Haß, & P. Storjohann (Eds.), *Handbuch Wort und Wortschatz* (pp. 274–294). Berlin: de Gruyter Mouton.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism.* Amsterdam: John Benjamins.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 121–134). Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second language.* Amsterdam: John Benjamins.

Pavlenko, A. (2004). L2 influence and L1 attrition in adult bilingualism. In M. S. Schmid, B. Köpke, M. Kejzer, & L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 47–59). Amsterdam: John Benjamins.

Pavlenko, A. (2009). *The bilingual mental lexicon: Interdisciplinary approaches*. Bristol: Multilingual Matters.

Pavlenko, A. (2011). *Thinking and speaking in two languages*. Clevedon: Multilingual matters.

Perkuhn, R., Belica, C., Kupietz, M., Keibel, H., & Hennig, S. (2009). *DeReWo: Korpusbasierte Wortformenliste*. Mannheim: Institut für Deutsche Sprache.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology, 22*(3), 415–422.

Potter, M. C., So, K., Eckardt, B. V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23–38.

Prior, A., Macwhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods, 39*(4), 1029–1038.

Prior, A., Kroll, J. F., & MacWhinney, B. (2013). Translation ambiguity but not word class predicts translation performance. *Bilingualism: Language and Cognition, 13,* 253–262.

Rast, R. (2008). *Foreign language input: Initial processing*. Clevedon: Multilingual Matters.

Ringbom, H. (1983). Borrowing and lexical transfer. *Applied Linguistics, 4*(3), 207–212.

Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.

Ringbom, H. (2001). Lexical transfer in L3 production. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 59–67). Clevedon: Multilingual Matters.

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters.

Roehr-Brackin, K. & Tellier, A. (2019). The role of language-analytic ability in children's instructed second language learning. *Studies in Second Language Acquisition, 41*(5), 111–1131.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). Oxford: Academic Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–40). Hillsdale, NJ: Lawrence Erlbaum.

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition, 18,* 179–190.

Rothman, J., & Cabrelli Amaro, J. (2010). What variables condition syntactic transfer? A look at the L3 initial state. *Second Language Research, 26*(2), 189–218.

Rothman, J., Gonzales Alonso, J., & Puig-Mayenco, E. (2019). *Third language acquisition and linguistic transfer.* Cambridge: Cambridge University Press.

Runnqvist, E. & Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? *Bilingualism: Language and Cognition, 15*(2), 365–377.

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition, 13*(2), 231–252.

Sayehli, S. (2013). *Developmental perspectives on transfer in third language acquisition*. Unpublished doctoral dissertation. Lund University, Sweden.

Schachter, J. (1974). An error in error analysis. *Language Learning*, *24*(2), 205–214.

Schmid, H.-J., (2007). Entrenchment, salience, and basic levels. In D. Geeraerts, & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics* (pp. 117-138). Oxford: Oxford University Press.

Schmid, M., & Köpke, B. (2008). L1 attrition and the mental lexicon. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 209–238). Bristol: Multilingual Matters.

Schmid, M., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition, 17*(4), 729–748.

Schmid, M., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism, 7*(6), 637–667.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Teaching (IRAL), 10,* 209–241.

Selinker, L. (1992). *Rediscovering interlanguage.* London: Longman.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16,* 155–172.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Sharwood Smith, M. (1989). Crosslinguistic influence in language loss. In K. Hyltenstam, & L. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss* (pp. 185-201). Cambridge: Cambridge University Press.

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

Singleton, D. (1987). Mother- and other-tongue influence on learner French. *Studies in Second Language Acquisition, 9,* 327–346.

Singleton, D. (1999). *Exploring the second language lexicon.* Cambridge: Cambridge University Press.

Singleton, D., & O'Laoire, M. (2006). Psychotypology and the 'L2 factor' in cross-lexical interaction: An analysis of English and Irish influence in learner French. In M. Bendtsen, M. Björklund, C. Fant, & L. Forsman (Eds.), *Språk, lärande och utbildning i sikte* (pp. 191–205). Vaasa: Faculty of Education, Åbo Akademi.

Slabakova, R. (2016). The Scalpel Model of third language acquisition. *International Journal of Bilingualism, 21*(6), 651–665.

Slobin, D. (1993). Adult language acquisition: A view from child language study. In C. Perdue (Ed.), *Adult language acquisition: cross-linguistic perspectives* (pp. 239–252). Cambridge: Cambridge University Press.

Stedje, A. (1977). Tredjespråksinterferens i fritt tal - en jämförande studie. In R. Palmberg, & H. Ringbom (Eds.), *Föredrag vid konferensen om kontrastiv lingvistik och felanalys*, (pp. 141–158). Turku: Abo Akademi.

Steinhauer, K., Panda, E., & Drury, J. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related potentials. *Neurological Correlates of Second Language Learning in Adults, 25*(1), 13–41.

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31.

Suhonen, L. (2015). *Conceptual cross-linguistic influence: Exploring the L1 lemma mediation hypothesis in L3 vocabulary acquisition* (Unpublished master's thesis). Lund University.

Talmy, L. (1975). Figure and ground in complex sentences. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society,* 419–430.

Taylor, J. (2003). *Linguistic categorization*. Oxford: Oxford University Press.

Thorndike, E. L. (1923). The influence of first year Latin upon the availability to read English. *School Sociology, 17,* 165–168.

Tokowicz, N. (2014). *Lexical processing and second language acquisition.* New York: Routledge.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness. Consequences of ambiguity within and across languages. *Language and Cognitive Processes, 22*, 727–779.

Tokowicz, N., Kroll, J. F., De Groot, A. M. B., & van Hell, J. G. (2002). Number of translation norms for Dutch–English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, and Computers, 34*, 435–451.

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics 4*(3), 273–290.

Ullman, M. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2*(10), 717–726

Ullman, M. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition. Methods, theory, and practice* (pp. 141–178). Washington: Georgetown University Press.

van Hell, J. G., & De Groot, A. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *The Quarterly Journal of Experimental Psychology, 51*(A), 41–63.

van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin and Review, 9*, 780–789.

van Hell, J. G., & Mahn, A. C. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and in experienced learners. *Language Learning, 47*, 507–546.

van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*(3), 458–483.

Vildomec, V. (1963). *Multilingualism*. Netherlands: Sythoff-Leyden.

Volodina, E., & Johansson, S. (2012). Introducing Swedish Kelly-list, a new lexical e-resource for Swedish. *Proceedings of LREC 2012*. Istanbul: ELRA.

Wei, L. (2000). Dimensions of bilingualism. In L. Wei (Ed.), *The bilingualism reader*, (pp. 3–25). New York: Routledge.

Wei, L. (2006). The multilingual mental lexicon and lemma transfer in third language learning. *International Journal of Multilingualism*, *3*(2), 88–104.

Weinreich, U. (1953). *Languages in contact.* New York: Linguistic Circle of New York.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2016). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism, 21*(6), 666–682.

White, A., Malt, B. C., & Storms, G. (2017). Convergence in the bilingual lexicon: A pre-registered replication of previous studies. *Frontiers in Psychology*, *7*, 2081.

Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics, 19*(3), 295–333.

Xia, C. (2017). Psychotypology of Chinese learners of English and its influence on the acquisition of metaphorical expressions: An offline study. *Cambridge Occasional Papers in Linguistics, 10,* 237–255.

Yu, L., & Odlin, T. (2016). *New perspectives on transfer in second language learning*. Clevedon: Multilingual matters.

Appendices

Appendix A Study 1-2: Background questionnaire 214
Appendix B Study 1: Sleep questionnaire 216
Appendix C Study 1: Form-meaning pairs 217
Appendix D Study 1: Primes and targets 219
Appendix E Study 1: Additional information 223
Appendix F Study 2: Form-meaning pairs 234
Appendix G Study 2: Primes and targets 236
Appendix H Study 2: E-CLSA Statements 240
Appendix I Study 2: Additional information 242
Appendix J Study 3: Background questionnaire (Long) 249
Appendix K Study 3: Background questionnaire (Short) 255
Appendix L Study 3-4: Stimuli creation process 258
Appendix M Study 3-4: Stimuli 260
Appendix N Study 3: Proficiency measures in Swedish 267

Appendix A Study 1-2: Background questionnaire

Answers in this form are analyzed anonymously

_________________ participant identifier __________ date ________ revision

Should you feel that the answer options for a particular question are not representative of your situation, you can include additional details at the end of this questionnaire.

1. How old are you? I am ____ years old

2. I am [ ] female [ ] male [ ] other *(tick the box next to your answer)*

3. Education (your current or most recent educational level, even if you have not finished the degree) *(if you are between degrees, mark the previous one) (tick the box next to your answer)*

[ ] Graduate school (PhD/MD/JD, or equivalent) [ ] High school

[ ] Graduate school (Masters, or equivalent) [ ] Middle school

[ ] College (BA, BS, or equivalent) [ ] Other (specify): ______________

4. Indicate your native language(s) and any other languages you have studied or learned, the age at which you started using each language in terms of listening, speaking, reading, and writing, and the total number of years you have spent using or learning each language.

Language	Mother tongue	Listening	Speaking	Reading	Writing	Years of Use
English	Yes [ ] No [ ]					
Swedish	Yes [ ] No [ ]					
___________	Yes [ ] No [ ]					
___________	Yes [ ] No [ ]					
___________	Yes [ ] No [ ]					

Example: if you started learning French at school at the age of 10 and stopped learning the language at the age of 12 and haven't used much of the language since, then indicate a total of two years of use.

5. Country of residence __________________ **Country of origin** __________________

If the country of residence and the country of origin are different, when did you move to the country where you currently live? month ______ year ______

6. Rate your language learning skill. In other words, how good do you feel you are at learning new languages, relative to your friends or other people you know? *(circle one)*

Very poor **1 2 3 4 5 6 7** Excellent

7. Do you like learning languages? *(circle one)*

Not at all **1 2 3 4 5 6 7** Very much

8. Is it easy for you to communicate with people from other countries? *(circle one)*

Very hard **1 2 3 4 5 6 7** Very easy

Answers in this form are analyzed anonymously

____________________ __________ ________
participant identifier date revision

9. If you have lived or travelled in countries other than your country of residence or country of origin for three or more months, then indicate the name of the country, your length of stay, the language used,[1] and the frequency of use of the language for each country.

Country	Length of stay[2] *(months)*	Language used	Frequency of use[4] *(circle one)*
			Never **1 2 3 4 5 6 7** Always
			Never **1 2 3 4 5 6 7** Always
			Never **1 2 3 4 5 6 7** Always
			Never **1 2 3 4 5 6 7** Always

1) Other than your mother tongue
2) You may have been to the country on multiple occasions, each for a different length of time. Add all the trips together.
3) If there is not enough space in the table above, continue on the last page
4) Please rate your use of the local language according to the following scale (circle the number in the table)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	*2*	*3*	*4*	*5*	*6*	*7*

10. Indicate the age at which you started using each of the languages you have studied or learned in the following environments.

	At home	With friends	At school	At work	Language software	Online games
English						
__________						
__________						
__________						

11. What language do you usually speak with your family? *(if more than one language, write all)*

____

12. Indicate the language used by your teachers for instruction at each educational level. *(If the instructional language switched during any educational level, then indicate the "Switched to" language)*

	Language	(Switched to)
Elementary school	__________________	__________________
Middle school	__________________	__________________
High school	__________________	__________________
College / university	__________________	__________________

13. You can also provide any other information about your language background or usage.

Appendix B Study 1: Sleep questionnaire

Answers in this form are analyzed anonymously

___________________	__________	________
participant identifier	date	revision

Should you feel that the answer options for a particular question are not representative of your situation, you can include additional details at the end of this questionnaire.

1. Rate the quality of sleep you had last night *(circle one)*

Very poor **1 2 3 4 5 6 7** Excellent

2. I had trouble falling asleep *(circle one)*

Not at all **1 2 3 4 5 6 7** Very much

3. Do you agree with the following statement: I felt like I did not get deep sleep *(circle one)*

Disagree **1 2 3 4 5 6 7** Agree

4. I was still tired even after waking up in the morning *(circle one)*

Disagree **1 2 3 4 5 6 7** Agree

5. Approximately how many hours of sleep did you have last night?

____

6. How many units of alcohol have you consumed between now and the first data collection session? *(One small beer or cider, a small glass of wine or one shot glass of hard liquor is roughly equivalent to one unit of alcohol. If you haven't consumed any alcohol write down zero.)*

____

7. Please comment below to indicate any additional answers to any of the questions above that you feel better describe the time between the first and second data collection sessions.

____

____

____

____

____

____

Appendix C Study 1: Form-meaning pairs

Kontu v1: Form-Meaning Pairs

1. Experimental Form-Meaning Pairs

	Form	Meaning 1	Meaning 2
Item	Koti	Flag	Ticket
Item	Talo	Snake	Queue
Item	Lappu	Bike	Wheel
Item	Pullo	Paint	Goal

2. Control Form-Meaning Pairs

Form	Structure	Meaning
Sota	CVCV	Crest
Lukko	CVCCV	Logo
Palo	CVCV	Banner
Kallo	CVCCV	Pin
Kelo	CVCV	Receipt
Talli	CVCCV	Note
Puro	CVCV	Prescription
Tuppi	CVCCV	Frog
Sulo	CVCV	Gekko
Matto	CVCCV	Crocodile
Jopo	CVCV	Turtle
Rakki	CVCCV	Line
Sato	CVCV	Pile
Koppa	CVCCV	Stack
Tuli	CVCV	Car
Nokka	CVCCV	Bus
Suti	CVCV	Skis
Suppo	CVCCV	Sledge
Huti	CVCV	Door
Sappi	CVCCV	Engine
Mopo	CVCV	Windshield

Form	Structure	Meaning
Kulma	CVCCV	Watercolor
Ropo	CVCV	Highlighter
Tulva	CVCCV	Nailpolish
Sini	CVCV	Lipstick
Halko	CVCCV	Basket
Suvi	CVCV	Target
Urpo	CVCCV	Hoop

Appendix D Study 1: Primes and targets

Table D1

Prime-target pairings in the distractor condition

	prime	TARGET
DISTRACTORS	bool	WROD
	hoop	WROS
	zelch	BANGIR
	target	BANGIT
	floof	WROST
	basket	WROMP
	bonh	FLEEP
	lipstick	FLEEC
	fleec	BOST
	nailpolish	BONH
	wromp	FLOOL
	highlighter	FLOOF
	bangit	ZENTH
	watercolor	ZELCH
	wros	BOAD
	windshield	BOOL

Note. The items in the distractor condition were non-words in English. Their purpose was to mask the actual purpose of the task from the participants.

Table D2

Prime-target pairings in the 'kontu' condition

	prime	TARGET
KONTU	flag	TICKET
	ticket	FLAG
	snake	QUEUE
	queue	SNAKE
	bike	WHEEL
	wheel	BIKE
	goal	PAINT
	paint	GOAL

Note. The items in the 'kontu' condition were items that had translation ambiguity between the L1 English and the L2 Kontu v1.

Table D3

Prime-target pairings in the baseline condition

	prime	TARGET
	logo	CREST
BASELINE	hobo	CREST
	banner	LOGO
	tunnel	LOGO
	pin	BANNER
	toe	BANNER
	crest	PIN
	toast	PIN
	note	RECEIPT
	tree	RECEIPT
	prescription	NOTE
	refrigerator	NOTE
	receipt	PRESCRIPTION
	trumpet	PRESCRIPTION
	gecko	FROG
	judge	FROG
	crocodile	GECKO
	paparazzi	GECKO
	turtle	CROCODILE
	eraser	CROCODILE
	frog	TURTLE
	roof	TURTLE
	pile	LINE
	book	LINE
	stack	PILE
	floor	PILE
	line	STACK
	tool	STACK
	bus	CAR
	tea	CAR
	skis	BUS
	flea	BUS
	sledge	SKIS
	butter	SKIS
	car	SLEDGE
	tar	SLEDGE
	engine	DOOR
	flower	DOOR
	windshield	ENGINE

	dishwasher	ENGINE
	door	WINDSHIELD
	lamp	WINDSHIELD
	highlighter	WATERCOLOR
	veterinarian	WATERCOLOR
	nailpolish	HIGHLIGHTER
	pillowcase	HIGHLIGHTER
	lipstick	NAILPOLISH
	mattress	NAILPOLISH
	watercolor	LIPSTICK
	television	LIPSTICK
	target	BASKET
	wall	BASKET
	hoop	TARGET
	wall	TARGET
	basket	HOOP
	artist	HOOP

Note. The items in the baseline condition were items that had no translation ambiguity between the L1 English and the L2 Kontu v1. Half of the prime-target conditions had a related prime (from the same category of words in the experiment) and half had an unrelated prime (matched in length and roughly in frequency).

Table D4

Prime-target pairings in the 'pseudo' condition

	prime	TARGET
	bami	CREST
PSEUDO	bodmir	LOGO
	bos	BANNER
	wrost	PIN
	bomp	RECEIPT
	plextraction	NOTE
	zeboans	PRESCRIPTION
	bodmi	FROG
	plelidest	GECKO
	bodha	CROCODILE
	wrob	TURTLE
	boft	LINE
	wroft	PILE
	bair	STACK
	bov	CAR
	wrov	BUS
	flymph	SKIS

boc	SLEDGE
mccott	DOOR
zertbriend	ENGINE
boub	WINDSHIELD
tophtrunctel	WATERCOLOR
tophrujuct	HIGHLIGHTER
bodthike	NAILPOLISH
mililipal	LIPSTICK
bautic	BASKET
bour	TARGET
bodmiz	HOOP
bangir	FLAG
wrod	TICKET
fleed	SNAKE
wrost	QUEUE
zenth	GOAL
flool	BIKE
boad	PAINT
flool	WHEEL

Note. The items in the 'pseudo' condition were items with a pseudoword as the prime and one of the experimental words as a target.

Data

Learning data

The time on task (ToT) data for the learning and assessment blocks contained a total of 22,386 observations of time on task and multiple choice responses.[60] The data was positively-skewed (M = 2565, Mdn = 1901, SD = 2656, skewness = 12.6 [SE = 0.0164], kurtosis = 504 [SE = 0.027]) and contained both unintended presses and unplanned breaks (range = 152324ms). Response times of <100ms were removed (n = 290) which represented 1.28% of the data as well as one instance of obvious unplanned break of 152 seconds. This left 22,095 time observations. The two conditions were almost equally affected by the missing data and the transformations with 97.42% data left in the single translation condition and 97.48% data left in the two translation condition.

Table E1

Distribution of time spent on task acquiring form-meaning mappings in Kontu v1

N	*M*	*Mdn*	*SD*	Range	Min	Max	Skewness	Kurtosis
22,095	2592	1921	2459	62696	103	62799	5.87	72.7

Note. Time on task (ms) for all recorded observations, i.e., both observations from learning tasks and assessment tasks across all participants and all items. Reported distribution is after trimming.

As we can see from Table E1 above, the resulting data is still non-normally distributed with skewness of 5.87 (SE = 0.0165) and kurtosis of 72.7 (SE = 0.0330). In the learning segments the participants were allowed to work through the learning sequences at their own pace which results in some of the items having taken the participants a long time. The primary purpose of the learning segments was to maximize the participants' learning of the items and as such strict time limits were not enforced.

To improve the distribution of the data, while keeping as much of the data as possible, observations beyond x (i.e., arbitrary number) standard deviations could

[60] Some 301 response time values (1.33%) were missing due to two instances of the presentation software temporarily freezing during data collection.

be removed. However, observations exceeding three standard deviations from the mean (10533ms) comprise exclusively of observations from the first five presentation instances of a particular item. Furthermore, they are predominantly from the first instance (46%). They seem to be caused by the participants occasionally spending extended amount of time on the first occurrence rather than by external factors. It seems that some of the participants clearly attempted a "get it right from the beginning" -type of learning strategy. While removing the aforementioned observations would account for only 1.2% of the data, it cannot be justified to remove them since their distribution is so heavily focused on the first instances.

Learning outcome data

In addition to ToT data from the learning sequences and learning assessment components, the data available on learning outcomes include binary data on the participants' LOT scores in the form of correct and incorrect responses. These 5,400 observations have been annotated with the available background data as well as the associated learning data from the 16 occurrences in the form of ToTs. Overall accuracy in the four learning outcome assessment tasks was extremely high, which indicates that the participants actively engaged in the learning module. No trimming or transformations were done to this data.

Priming data

The response time (RT) data for priming tasks (pre-, immediate post-, and delayed post-test) contained a total of 10,440 observations. The data was positively-skewed (M = 529, Mdn = 501, SD = 156, skewness = 7.36 [SE = 0.0240], kurtosis = 180 [SE = 0.0479]) and contained both unintended presses and unplanned breaks (range = 5837). Response times of <200ms and >1000ms were removed (n = 130) which represented 1.24% of the data. This left 10,310 RT observations. Distractor items were most affected (1,7%). Out of the items in the conditions of interest, 0.7% of baseline items and 0.2% of items related in Kontu v1 were affected. The data is still positively skewed, as can be seen from Table E2 below.

Table E2

Distribution of response time data in priming tasks

N	*M*	*Mdn*	*SD*	Range	Min	Max	Skewness	Kurtosis
10,310	520	499	114	785	214	999	1.06	1.54

Note. Response times (ms) include all recorded observations across all participants and all items. Reported distribution is after trimming.

Aptitude data

The aptitude data consists of three types of data: self-evaluations, working memory (WM) scores, and cognitive control (CC) scores. Neither trimming nor transformations were performed for self-evaluations. These have been added for both accuracy and RT observations as predictors. No trimming was deemed necessary for WM or CC data. Both data sets contained timed out responses (37 instances). For cognitive control, there were no response times of <100ms. The data contains 1,920 observations and is right-skewed (*M* = 611, *Mdn* = 579, *SD* = 171, skewness = 2.66 [*SE* = 0.0559], kurtosis = 11.1 [*SE* = 0.112]). For WM, there were two observations of response times of <100ms. The RT data contains 3,000 observations and is right-skewed for WM2 (*M* = 777, *Mdn* = 720, *SD* = 266, skewness = 1.18 [*SE* = 0.0636], kurtosis = 1.63 [*SE* = 0.127]) and closer to normal distribution for WM3 (*M* = 834, *Mdn* = 775, *SD* = 331, skewness = 0.833 [*SE* = 0.0635], kurtosis = 0.418 [*SE* = 0.127]). The 2-back task was reported in the debriefing by multiple participants to be quite easy in comparison to the 3-back task.

Results

Aptitude

Cognitive control (CC) and working memory (WM), here subsumed under *aptitude,* have been postulated to have a major impact on the acquisition and processing of vocabulary (Ellis & Sinclair, 1996). Three measures (each with multiple components) were used as independent variables: self-evaluated aptitude, performance in CC task, and performance in WM tasks. Despite having limited experience in language learning, the participants self-rated their aptitude relatively high on a seven-point Likert-scale; see Table E3 below. The table shows that the participants rate their attitude towards learning higher than their aptitude. There is a fair amount of variance in the data with responses ranging almost throughout the whole scale. In addition to self-ratings, the participants' aptitude skills were measured with respect to CC and WM. A flanker task was used for CC. The results are presented in Table E4 below.

Table E3

Self-ratings of language learning aptitude, attitude, and cross-cultural communicative skills

	N	*M*	*MDN*	*SD*	MIN	MAX
APTITUDE	30	4.13	4.00	1.11	2	6
ATTITUDE	30	5.07	5.00	1.17	2	7
COMMUNICATION	30	4.23	4.50	1.55	1	7

Note. Self-ratings on aptitude, attitude, and ability to communicate cross-culturally on a seven-point Likert-scale.

As expected, with a few exceptions, the participants were faster at the congruent items than the incongruent items. Accuracy in the task is also high (*M* = 0.927). One participant performed below chance despite on-screen feedback.

Table E4

Performance in the flanker-task

	N	*M*	*MDN*	*SD*	MIN	MAX
INCONGRUENT	960	621	621	72.6	454	832
CONGRUENT	960	601	614	68	473	747
FLANKER	30	20.0	15	41.3	-57.8	144
ACCURACY	960	0.927	0.953	0.097	0.484	1.00

Note. In incongruent items the distractors belonged to a different category than the focus item. In congruent items, both belonged to the same category. Time measures reported in milliseconds.

The participants' WM was tested by means of 2-back and 3-back tasks. Table E5 below presents the results. Expectedly, the 2-back task was easier than the 3-back task, which shows in both accuracy and RTs, particularly for the repeated trials. The items timed out at 2,000ms and very few observations could be classified as premature responses.

Table E5

Working memory scores for 2-back and 3-back tasks

	N	*M*	*MDN*	*SD*	MIN	MAX
2-BACK						
RT	1481	777	720	266	87	1992
ACCURACY	1500	0.911	1.00	0.284	0	1
REPETITION	149	0.680	1.00	0.468	0	1
3-BACK						
RT	1486	834	775	331	50	1996
ACCURACY	1500	0.854	1.00	0.353	0	1
REPETITION	178	0.461	0.00	0.500	0	1
2-BACK AVG						
RT	30	778	757	191	509	1329
ACCURACY	30	0.911	0.940	0.083	0.540	0.980
REPETITION	30	0.680	0.800	0.266	0.00	1.00
3-BACK AVG						
RT	30	837	845	249	442	1479
ACCURACY	30	0.855	0.860	0.071	0.600	0.980
REPETITION	30	0.461	0.417	0.242	0.167	1.00
OVERALL AVG						
RT	30	808	810	213	485	1404
ACCURACY	30	0.883	0.890	0.055	0.680	0.950
REPETITION	30	0.571	0.542	0.210	0.083	0.900

Note. 2-back and 3-back refer to the stimulus distance in the *n*-back task. AVG refers to average score per participants (n=30) that are used as fixed effects in the learning outcome and CLI analyses. Overall average refers to the means of the 2-back and 3-back tasks combined. RT is response time (0-2,000 as the tasks timed out at 2,000) in milliseconds, ACCURACY the ratio between correct and incorrect responses (0-1), and REPETITION the ratio between correct and incorrect responses in repeated stimulus (0-1). The 3-back task was more difficult than the 2-back task based on both response times, overall accuracy, and accuracy for repeated trials.

The self-rated aptitude measures correlate with each other with aptitude and attitude having the highest correlation (r = 0.738). Self-evaluations seem to correlate more with accuracy than with processing. The same applies for the aptitude measures, where the correlations between accuracy and RT measures across tasks are higher than for the two types within the same task. See Table E6 below for a correlation matrix. The aptitude measures that are used as predictors in the linear mixed effects models are flanker-effect (CCFLA) for cognitive control, and repeated trial accuracy in the 3-back task (W3AC) for working memory.

Table E6

Correlation matrix for aptitude measures

		APT	ATT	COM	CCFLA	CCAC	W2RT	W2AC	W3RT	W3AC
APT	*r*	–								
	p	–								
ATT	*r*	0.738	–							
	p	< .001	–							
COM	*r*	0.526	0.676	–						
	p	0.003	< .001	–						
CCFLA	*r*	0.255	0.236	0.276	–					
	p	0.174	0.210	0.140	–					
CCAC	*r*	0.389	0.519	0.416	0.341	–				
	p	0.034	0.003	0.022	0.066	–				
W2RT	*r*	-0.115	-0.028	0.205	-0.292	-0.165	–			
	p	0.547	0.884	0.278	0.117	0.385	–			
W2AC	*r*	0.150	0.004	-0.131	-0.059	0.148	0.123	–		
	p	0.428	0.981	0.491	0.757	0.435	0.517	–		
W3RT	*r*	-0.111	-0.178	0.057	-0.337	-0.357	0.875	0.331	–	
	p	0.559	0.347	0.763	0.069	0.053	< .001	0.074	–	
W3AC	*r*	0.406	0.414	0.347	0.361	0.238	0.048	0.371	0.057	–
	p	0.026	0.023	0.060	0.050	0.205	0.800	0.044	0.764	–

Note. Aptitude (APT), attitude (ATT), and intercultural communicative ability (COM) are self-rated. Cognitive control are represented by two measures: flanker effect (CCFLA) and accuracy in the flanker task (CCAC). W2 and W3 refer to the 2-back and 3-back tasks for working memory.

Forward CLI

Table E7 below shows the ToTs for the two types of stimuli available in the learning task throughout their sixteen occurrences. The presentation order was randomized and data below is based on the tagged order of occurrence. The data presented below shows that the ToTs reduce successively from the first occurrences towards the last one.

When comparing the first and second instantiations of the same form (with different meaning), the second instantiation is longer on average. This is presented in Table E8 below that represents a linear mixed effects model (LME). PARTICIPANT and ITEM have been set as random intercepts. The only fixed effect is a binary variable of whether the item is presented in its first form instantiation or its second form instantiation with a different meaning. The effect is not significant ($p = 0.71$), meaning that the analysis does not support the hypothesis.

Furthermore, a LME was performed to model the development of ToTs during learning between first and second instantiations of the same form in assessment data. Random intercepts were set for PARTICIPANT and ITEM. The only fixed effect is INSTANCE, i.e., whether the particular form was presented for the first or the second time. There is an estimated difference of 674.6ms but this effect is not significant ($p = 0.101$). Table E9 below represents the LME.

Table E7

Time spent on task (learning sequences) for acquiring form-meaning mappings in Kontu v1

OCCURRENCE	ITEM TYPE	MEAN	*SD*	MEDIAN
1	MONO	6171	4760	4957
	POLY	6777	5924	5150
2	MONO	5349	3985	4378
	POLY	5303	3697	4322
3	MONO	4785	3175	4039
	POLY	4629	3211	3799
4	MONO	4191	2480	3706
TEST 1	POLY	4216	3138	3502
5	MONO	2322	1299	2064
	POLY	2641	4040	2091
6	MONO	2028	1039	1820
	POLY	1983	1055	1713
7	MONO	1926	1022	1709
	POLY	1821	998	1559
8	MONO	1760	928	1567
TEST 2	POLY	1720	937	1490
9	MONO	1787	1059	1555
	POLY	1732	970	1551
10	MONO	1624	901	1436
	POLY	1670	1005	1464
11	MONO	1508	890	1281
	POLY	1566	850	1343
12	MONO	1439	813	1206
TEST 3	POLY	1439	816	1216
13	MONO	1591	980	1395
	POLY	1503	773	1319
14	MONO	1415	759	1238
	POLY	1433	781	1238
15	MONO	1400	810	1196
	POLY	1379	738	1165
16	MONO	1358	757	1152
TEST 4	POLY	1475	825	1259

Note. Occurrence is the amount of times a form-meaning pair has been presented. Item type refers to one or two translation equivalents in the L1. All presented times are in milliseconds across participants and items.

Table E8

First instance vs. second instance of repeated form with different meaning in learning

Scaled residuals

MIN	1Q	MEDIAN	3Q	MAX
-2.1367	-0.3545	-0.1172	0.1368	9.2365

Random effects

GROUPS	NAME	VARIANCE	SE
ITEM	INTERCEPT	890638	9.2365
PARTICIPANT	INTERCEPT	20504766	4528.2
RESIDUAL		9431200	3071.0

Number of obs: 224, participant, 28; item, 32

Fixed effects

	ESTIMATE	SE	df	t	*p*	
INTERCEPT	6483.59	1205.56	80.33	5.378	< .001	***
INSTANCE	194.98	524.03	192.32	0.372	0.71	

REML criterion at convergence 4380.8

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to the first instance of a particular Kontu v1 form irrespective of meaning. INSTANCE refers to the first instantiation of a second meaning for the same Kontu v1 form. The difference between a first instance of a form and a second instance of a form but with another meaning is not significant. Formula: ToT ~ instance + (1|participant) + (1|item).

Table E9

First instance vs. second instance of repeated form with different meaning in assessment

Scaled residuals

	MIN	1Q	MEDIAN	3Q	MAX
	-1.4802	-0.4994	-0.2519	0.2408	5.7658

Random effects

GROUPS	NAME	VARIANCE	SE
PARTICIPANT	INTERCEPT	2409146	1552.1
ITEM	INTERCEPT	240273	490.2
RESIDUAL		9431200	3071.0

Number of obs: 234, participant, 30; item, 8

Fixed effects

	ESTIMATE	SE	df	t	p	
INTERCEPT	3594.8	721.7	129.3	4.981	< .001	***
INSTANCE	674.6	409.7	202.4	1.647	0.101	

REML criterion at convergence 4429.3

Note. Linear mixed model fit by REML. T-tests use Satterthwaite's method ['lmerModLmerTest']. Performed in R version 3.6.0 using lme4 (1.1-23) and lmerTest (3.1-2). INTERCEPT is set to the first instance of a particular Kontu v1 form irrespective of meaning. INSTANCE refers to the first instantiation of a second meaning for the same Kontu v1 form. The difference between a first instance and a first instance with another meaning is not significant. Formula: ToT ~ instance + (1|participant) + (1|item).

Reverse CLI

Table E10 below presents RTs in the pre-test, post-test, and delayed post-test priming tasks. These were performed solely in the L1. No adjustments to generic learning effects have been made in the table below.

Table E10

Response times in the priming task by item type

TEST	ITEM TYPE	MEAN	ST. DEV.	MEDIAN
PRE-TEST	BASELINE	510	110	489
	DISTRACTOR	652	146	643
	KONTU	492	92.6	475
	PSEUDO	533	129	505
POST-TEST	BASELINE	512	101	495
	DISTRACTOR	647	105	634
	KONTU	503	94.2	487
	PSEUDO	535	112	516
DELAYED POST-TEST	BASELINE	488	97.5	472
	DISTRACTOR	597	109	593
	KONTU	472	79.5	466
	PSEUDO	514	108	498

Note. RTs in priming task. BASELINE to refers to baseline items that are translation equivalents of Kontu words with a single translation equivalent in the L1. DISTRACTOR refers to nonce words that are legal in English phonotactics. KONTU refers to item pairs that are related in Kontu but not in English. PSEUDO refers to English words that are preceded by pseudowords in the mask.

Appendix F Study 2: Form-meaning pairs

Kontu v2: Form-Meaning Pairs

1. Experimental Form-Meaning Pairs

	Form	Meaning 1	Meaning 2
Item	Koti	Flag	Ticket
Item	Talo	Snake	Queue
Item	Lappu	Bike	Wheel
Item	Pullo	Road	Goal

2. Control Form-Meaning Pairs

Form	Structure	Meaning
Sota	CVCV	Crest
Lukko	CVCCV	Logo
Palo	CVCV	Banner
Kallo	CVCCV	Pin
Kelo	CVCV	Receipt
Talli	CVCCV	Note
Puro	CVCV	Prescription
Tuppi	CVCCV	Frog
Sulo	CVCV	Gekko
Matto	CVCCV	Crocodile
Jopo	CVCV	Turtle
Rakki	CVCCV	Line
Sato	CVCV	Pile
Koppa	CVCCV	Stack
Tuli	CVCV	Car
Nokka	CVCCV	Bus
Suti	CVCV	Skis
Suppo	CVCCV	Sledge
Huti	CVCV	Door
Sappi	CVCCV	Engine
Mopo	CVCV	Windshield

Form	Structure	Meaning
Kulma	CVCCV	Watercolor
Ropo	CVCV	Highlighter
Tulva	CVCCV	Nailpolish
Sini	CVCV	Lipstick
Halko	CVCCV	Basket
Suvi	CVCV	Target
Urpo	CVCCV	Hoop

Appendix G Study 2: Primes and targets

Table G1

Prime-target pairings in the distractor condition

	prime (en)	TARGET (en)	prime (se)	TARGET (se)
	bool	WROD	nex	PLABBA
DISTRACTORS	hoop	WROS	målring	PLABDA
	zelch	BANGIR	näz	SILJETT
	target	BANGIT	skottavla	SILJEDD
	floof	WROST	mjux	ERN
	basket	WROMP	korg	ERD
	bonh	FLEEP	setez	JÖ
	lipstick	FLEEC	läppstift	JB
	fleec	BOST	jb	SETEL
	nailpolish	BONH	nagellack	SETEZ
	wromp	FLOOL	erd	MJUL
	highlighter	FLOOF	strykningspenna	MJUX
	bangit	ZENTH	siljedd	NÄG
	watercolor	ZELCH	vattenfärg	NÄZ
	wros	BOAD	plabda	NEL
	windshield	BOOL	vindruta	NEX

Note. The items in the distractor condition were non-words in English. Their purpose was to mask the actual purpose of the task from the participants.

Table G2

Prime-target pairings in the 'kontu' condition

	prime (en)	TARGET (en)	prime (se)	TARGET (se)
	flag	TICKET	flagga	BILJETT
KONTU	ticket	FLAG	biljett	FLAGGA
	snake	QUEUE	orm	KÖ
	queue	SNAKE	kö	ORM
	bike	WHEEL	cykel	HJUL
	wheel	BIKE	hjul	CYKEL
	goal	ROAD	väg	MÅL
	road	GOAL	mål	VÄG

Note. The items in the 'kontu' condition were items that had translation ambiguity between the L1 English and the L2 Kontu v1.

Table G3

Prime-target pairings in the baseline condition

	prime	TARGET	prime (se)	TARGET (se)
	logo	CREST	logga	KRÖN
BASELINE	hobo	CREST	matta	KRÖN
	banner	LOGO	baner	LOGGA
	tunnel	LOGO	taler	LOGGA
	pin	BANNER	pins	BANER
	toe	BANNER	höns	BANER
	crest	PIN	krön	PINS
	toast	PIN	bröd	PINS
	note	RECEIPT	nota	KVITTO
	tree	RECEIPT	träd	KVITTO
	prescription	NOTE	recept	ANTECKNING
	refrigerator	NOTE	toalett	ANTECKNING
	receipt	PRESCRIPTION	recept	RECEPT
	trumpet	PRESCRIPTION	toalett	RECEPT
	gecko	FROG	gekko	GRODA
	judge	FROG	matta	GRODA
	crocodile	GECKO	krokodil	GECKO
	paparazzi	GECKO	paparazzi	GECKO
	turtle	CROCODILE	sköldpadda	KROKODIL
	eraser	CROCODILE	suddgummi	KROKODIL
	frog	TURTLE	groda	SKÖLDPADDA
	roof	TURTLE	tak	SKÖLDPADDA
	pile	LINE	pil	LINJE
	book	LINE	bok	LINJE
	stack	PILE	staple	PIL
	floor	PILE	golv	PIL
	line	STACK	linje	STAPEL
	tool	STACK	verktyg	STAPEL
	bus	CAR	buss	BIL
	tea	CAR	te	BIL
	skis	BUS	skidor	BUSS
	flea	BUS	loppa	BUSS
	sledge	SKIS	kälke	SKIDOR
	butter	SKIS	flymph	SKIDOR
	car	SLEDGE	bil	KÄLKE
	tar	SLEDGE	tjär	KÄLKE
	engine	DOOR	motor	DÖRR
	flower	DOOR	blomma	DÖRR
	windshield	ENGINE	vindruta	MOTOR

dishwasher	ENGINE	diskmaskin	MOTOR
door	WINDSHIELD	dörr	VINDRUTA
lamp	WINDSHIELD	lampa	VINDRUTA
highlighter	WATERCOLOR	strykningspenna	VATTENFÄRG
veterinarian	WATERCOLOR	veterinär	VATTENFÄRG
nailpolish	HIGHLIGHTER	nagellack	STRYKNINGSPENNA
pillowcase	HIGHLIGHTER	örngott	STRYKNINGSPENNA
lipstick	NAILPOLISH	läppstift	NAGELLACK
mattress	NAILPOLISH	madrass	NAGELLACK
watercolor	LIPSTICK	vattenfärg	LÄPPSTIFT
television	LIPSTICK	television	LÄPPSTIFT
target	BASKET	skottavla	KORG
wall	BASKET	vän	KORG
hoop	TARGET	målring	SKOTTAVLA
wall	TARGET	vägg	SKOTTAVLA
basket	HOOP	korg	MÅLRING
artist	HOOP	vägg	MÅLRING

Note. The items in the baseline condition were items that had no translation ambiguity between the L1 English and the L2 Kontu v1. Half of the prime-target conditions had a related prime (from the same category of words in the experiment) and half had an unrelated prime (matched in length and roughly in frequency).

Table G4

Prime-target pairings in the 'pseudo' condition

	prime (en)	TARGET (en)	prime (se)	TARGET (se)
	bami	CREST	bammi	KRÖN
PSEUDO	bodmir	LOGO	bomir	LOGGA
	bos	BANNER	bons	BANER
	wrost	PIN	dern	PINS
	bomp	RECEIPT	boma	KVITTO
	plextraction	NOTE	deceft	ANTECKNING
	zeboans	PRESCRIPTION	deceft	RECEPT
	bodmi	FROG	bodmi	GRODA
	plelidest	GECKO	plelesit	GECKO
	bodha	CROCODILE	bodha	KROKODIL
	wrob	TURTLE	wrob	SKÖLDPADDA
	boft	LINE	bof	LINJE
	wroft	PILE	wroft	PIL
	bair	STACK	bair	STAPEL
	bov	CAR	bovt	BIL
	wrov	BUS	vrov	BUSS
	flymph	SKIS	flymph	SKIDOR
	boc	SLEDGE	boc	KÄLKE
	mccott	DOOR	mccott	DÖRR
	zertbriend	ENGINE	zertbried	MOTOR
	boub	WINDSHIELD	boub	VINDRUTA
	tophtrunctel	WATERCOLOR	tophtrunctel	VATTENFÄRG
	tophrujuct	HIGHLIGHTER	tophrujuct	STRYKNINGSPENNA
	bodthike	NAILPOLISH	bodthike	NAGELLACK
	mililipal	LIPSTICK	milipal	LÄPPSTIFT
	bautic	BASKET	bautic	KORG
	bour	TARGET	bour	SKOTTAVLA
	bodmiz	HOOP	bodmiz	MÅLRING
	bangir	FLAG	siljett	FLAGGA
	wrod	TICKET	plabba	BILJETT
	fleed	SNAKE	jö	ORM
	wrost	QUEUE	plabda	KÖ
	zenth	GOAL	mjul	CYKEL
	flool	BIKE	ljul	HJUL
	boad	ROAD	nel	VÄG
	flool	WHEEL	näg	MÅL

Note. The items in the 'pseudo' condition were items with a pseudoword as the prime and one of the experimental words as a target.

Appendix H Study 2: E-CLSA Statements

Overview

The following 12 statements were used in the E-CLSA. The language combinations depended on the study. Then using the E-CLSA with two languages, the placeholders Lx and Ly can be replaced with the languages in question resulting in 12 statements. With three languages, three versions of each statements are created combining all three combinations of the languages resulting in 36 statements. The headline represents the primary area of language knowledge probed in the statements. If using the following statements in future research, consider reversing the order of negative and positive statements as well as the order of presentation of the languages for a portion of your sample. Note that the statements below were created for a study with lexical knowledge as the outcome variable meaning that references to particularly vocabulary in statements of other areas of knowledge (e.g., phonology and spelling) are overproportionate.

Global

I think *Lx* and *Ly* are very similar to each other

If you know *Lx*, learning *Ly* is easy

Grammar

Lx grammar is a lot like *Ly* grammar

Idioms

Expressions in *Lx* can often be translated to *Ly* word-for-word

Phonology and pronunciation

Words in *Lx* are often pronounced in a similar way to words in *Ly*

I generally believe that the pronunciation of *Lx* is closely related to pronunciation of *Ly*

I generally believe that there is not much resemblance between the pronunciation system of *Lx* and that of *Ly*

I generally believe that *Lx* does not sound like *Ly*

Spelling

I see little similarity between *Lx* words and *Ly* words

Lx and *Ly* spelling are similar

Lx and *Ly* words do not look the same

Vocabulary

Lx and *Ly* have a lot of similar words

Appendix I Study 2: Additional information

Data

Learning data

The learning data was positively-skewed (*M* = 2460, *Mdn* = 1888, *SD* = 2233, skewness = 5.16 [*SE* = 0.0168], kurtosis = 54.6 [*SE* = 0.0336] and contained both unintended presses and unintended breaks (range = 49,075). ToTs of <100ms were removed (*n* = 363) which represented 1.68% of the data. This left 20,913 observations, presented in Table I1 below. The two conditions were almost equally affected by the missing data and the transformations with 96.85% data left in the single translation condition and 96.73% data left in the two translation condition.

Table I1

Time spent on task acquiring form-meaning mappings in Kontu v2

N	*M*	*Mdn*	*SD*	Range	Min	Max	Skewness	Kurtosis
20,913	2503	1918	2228	48971	104	49075	5.25	55.7

Note. ToTs (ms) include all recorded observations, i.e., both observations from learning tasks and assessment tasks across all participants and all items.

The resulting data is still non-normally distributed with skewness 5.25 (*SE* = 0.0169) and kurtosis of 55.7 (*SE* = 0.0339).

Priming data

The priming data was positively-skewed (*M* = 596, *Mdn* = 554, *SD* = 194, skewness = 4.27 [*SE* = 0.0209], kurtosis = 43.9 [*SE* = 0.0419] and contained both unintended presses and unplanned breaks (range 4156). The participants in Study 2 were overall slower in the priming task than the participants in Study1, which is expected given that they did the priming tasks both in their mother tongue as well as in their second language. For this reason, a higher upper trim point was chosen. Response times of <200ms and >1,400ms were removed (*n* = 98) which represented 0.70% of the data. The upper trimming point was set higher in Study 2 since the participants evaluated items in their L2. Distractor items were most affected (1.4%). The items used to analyze reverse CLI were almost equally affected at 0.67% for baseline and 0.52% for items with translation ambiguity. This left 13,590 observations, presented in Table I2 below.

Table I2

Distribution of response time data for priming tasks in Swedish and English

N	*M*	*Mdn*	*SD*	Range	Min	Max	Skewness	Kurtosis
13,590	587	553	155	1185	211	1396	1.50	3.39

Note. Response times (ms) include all recorded observations across all participants and all items.

Aptitude data

Upper spectrum trimming was not necessary for the RT data as both the WM and CC tasks timed out at 2,000ms (n = 185). RTs <100ms were removed from both WM and CC scores (n = 211). After trimming, there were 1,847 observations for CC and 2,598 for WM. The CC data is positively skewed (M = 659, Mdn = 624, SD = 193, skewness = 2.40 [SE = 0.0569], kurtosis = 9.30 [SE = 0.114]). For working memory, the RT data is positively skewed for both WM2 (M = 784, Mdn = 727, SD = 273, skewness = 1.05 [SE = 0.683], kurtosis = 1.38 [SE = 0.137]) and less skewed for WM3 (M = 817, Mdn = 792, SD = 287, skewness = 0.559 [SE = 0.0674], kurtosis = 0.895 [SE = 0.135]). Both WM and CC data were normalized using z-transformation for the linear mixed effect models.

Psychotypology data

The psychotypology data consists of two sets of data: 1) 1,080 measures of perceived language similarity between Swedish, English, and Finnish on a seven-point Likert scale, and 2) 1,680 RT measures on familiar items, Kontu v2 non-words, and novel lexical items in Kontu v2 based on the two potential source languages. These were computed into a participant-specific explicit and implicit psychotypology measures. The explicit measure was computed by first transforming negative and positive statement responses to a scale where higher always means more similar and then deducting the mean of Swedish-Finnish similarity evaluations (1-7) from the mean of English-Finnish similarity evaluations (1-7), resulting in a score of -6 to 6 where 0 is a perfectly balanced score. For the implicit measure, in line with the priming data, response times <200ms and >1,400ms were removed representing 2.56% of the data. The remaining data was moderately right-skewed (*M* = 656, *Mdn* = 626, *SD* = 205, skewness = 0.866 [*SE* = 0.0605], kurtosis = 0.737 [*SE* = 0.121]). The final implicit measure was computed by deducting the mean of the Swedish-based items from the mean of the English-based items resulting in a score where 0 represents a perfectly balanced score.

Results

Aptitude

Compared to the participants in Study 1, the participants in the present study rated their aptitude, attitude, and ability to communicate cross-culturally overall higher. The self-assessments are presented in Table I3 below.

Table I3

Self-ratings of language learning aptitude, attitude, and cross-cultural communicative skills

	N	*M*	*MDN*	*SD*	MIN	MAX
APTITUDE	30	4.80	5.00	0.77	3	6
ATTITUDE	30	5.73	6.00	1.31	2	7
COMMUNICATION	30	5.53	5.00	0.86	4	7

Note. The participants self-rated their aptitude, attitude, and ability to communicate cross-culturally on a seven-point Likert-scale. Attitude has the highest mean and aptitude the lowest.

The participants' CC was measured using a flanker task. The results are presented in Table I4 below. Accuracy in the CC task was high (M = 0.913). Two participants had 100% accuracy while one participant performed at chance.

Table I4

Performance in flanker-task (ms)

	N	*M*	*MDN*	*SD*	MIN	MAX
INCONGRUENT	928	670	652	91.7	509	900
CONGRUENT	928	649	647	82.6	522	911
FLANKER	29	20.5	24.1	42.1	-69.5	101
ACCURACY	928	0.913	0.953	0.104	0.500	1.00

Note. In incongruent items the distractors belonged to a different category than the focus item. In congruent items, both belonged to the same category. Time measures reported in milliseconds.

The WM of the participants was measured using a *n*-back task. Data for 2-back and 3-back versions was collected. The 2-back task was easier than the 3-back task judging from both mean accuracy and mean RTs. The items timed out at 2,000ms and RTs <100ms have been removed. Like in Study 1, the WM3 scores were used for the analysis and are presented in Table I5 below. One participant performed below chance in accuracy and the best performing participant had 94% overall accuracy in WM3.

Table I5

Working memory scores for 3-back tasks

	N	*M*	*MDN*	*SD*	MIN	MAX
RT	1316	817	729	287	112	1995
ACCURACY	1400	0.807		0.395	*binary*	
REPETITION	168	0.429		0.496	*binary*	

Note. RT is response time (0-2,000 as the tasks timed out at 2,000) in milliseconds, ACCURACY the ratio between correct and incorrect responses (0-1), and REPETITION the ratio between correct and incorrect responses in repeated stimulus (0-1).

Forward CLI

The development of ToTs in learning was tracked across item types over the course of learning. This represents an estimation of difficulty of acquiring a particular form-meaning mapping. Table I6 below contains the time spent on task for the two types of stimuli (with / without translation ambiguity) over their sixteen occurrences. Like in Study 1, presentation order was randomized and the data below is based on tagged order of presentation. In line with Study 1, the ToTs reduce over the course of learning.

Table I6

Time spent on task (learning sequences) for acquiring form-meaning mappings in Kontu v2

OCCURRENCE	ITEM TYPE	MEAN	*SD*	MEDIAN
1	MONO	5071	2931	4380
	POLY	5348	3383	4441
2	MONO	4787	3490	4064
	POLY	5093	4163	4025
3	MONO	4308	3289	3763
	POLY	4672	4335	3633
4	MONO	4037	3374	3577
TEST 1	POLY	4416	4980	3585
5	MONO	2145	1725	1796
	POLY	2128	1409	1742
6	MONO	1809	920	1620
	POLY	1882	1042	1681
7	MONO	1784	866	1621
	POLY	1701	821	1572
8	MONO	1757	851	1600
TEST 2	POLY	1723	837	1631
9	MONO	1731	892	1554
	POLY	1731	980	1575
10	MONO	1672	924	1456
	POLY	1698	912	1553
11	MONO	1602	923	1362
	POLY	1591	920	1395
12	MONO	1551	808	1429
TEST 3	POLY	1526	803	1350
13	MONO	1599	933	1376
	POLY	1670	967	1461
14	MONO	1532	835	1374
	POLY	1656	1086	1412
15	MONO	1468	833	1266
	POLY	1461	828	1278
16	MONO	1386	760	1231
TEST 4	POLY	1406	774	1211

Note. Occurrence is the amount of times a form-meaning pair has been presented. Item type refers to one or two translation equivalents in the L1. All times in ms across participants and items.

Reverse CLI

Table I7

Response times (ms) in the priming task by item type

	ITEM TYPE	MEAN	ST. DEV.	MEDIAN
SWEDISH PRE-TEST	BASELINE	582	151	546
	DISTRACTOR	741	163	705
	KONTU	554	139	529
	PSEUDO	601	163	559
POST-TEST	BASELINE	544	135	518
	DISTRACTOR	679	151	658
	KONTU	528	121	504
	PSEUDO	567	141	538
ENGLISH PRE-TEST	BASELINE	595	155	563
	DISTRACTOR	755	189	742
	KONTU	560	119	543
	PSEUDO	621	179	579
POST-TEST	BASELINE	560	125	540
	DISTRACTOR	710	159	695
	KONTU	541	115	513
	PSEUDO	585	148	557

Note. BASELINE to refers to baseline items that are translation equivalents of Kontu words with a single translation equivalent in the L1. DISTRACTOR refers to nonce words that are legal in English phonotactics. KONTU refers to item pairs that are related in Kontu but not in English. PSEUDO refers to English words that are preceded by pseudowords in the mask.

Appendix J Study 3: Background questionnaire (Long)

Answers in this form are analyzed anonymously

________________ participant identifier ________ date ______ revision

Should you feel that the answer options for a particular question are not representative of your situation, you can include additional details at the end of this questionnaire.

1. How old are you? I am ____ years old

2. I am [ ] female [ ] male [ ] other *(tick the box next to your answer)*

3. Education (your current or most recent educational level, even if you have not finished the degree) *(if you are between degrees, mark the previous one) (tick the box next to your answer)*

[ ] Graduate school (PhD/MD/JD, or equivalent)
[ ] Graduate school (Masters, or equivalent)
[ ] College (BA, BS, or equivalent)
[ ] High school
[ ] Middle school
[ ] Other (specify): ______________

4. Indicate your native language(s) and any other languages you have studied or learned, the age at which you started using each language in terms of listening, speaking, reading, and writing, and the total number of years you have spent using or learning each language.

Language	Mother tongue	Listening	Speaking	Reading	Writing	Years of Use
German	Yes [ ] No [ ]					
English	Yes [ ] No [ ]					
Swedish	Yes [ ] No [ ]					
__________	Yes [ ] No [ ]					
__________	Yes [ ] No [ ]					
__________	Yes [ ] No [ ]					

Example: if you started learning French at school at the age of 10 and stopped learning the language at the age of 12 and haven't used much of the language since, then indicate a total of two years of use.

5. Country of residence __________________ **Country of origin** __________________

If the country of residence and the country of origin are different, when did you move to the country where you currently live? month ______ year ______

6. Rate your language learning skill. In other words, how good do you feel you are at learning new languages, relative to your friends or other people you know? *(circle one)*

Very poor **1 2 3 4 5 6 7** Excellent

7. Do you like learning languages? *(circle one)*

Not at all **1 2 3 4 5 6 7** Very much

8. Is it easy for you to communicate with people from other countries? *(circle one)*

Very hard **1 2 3 4 5 6 7** Very easy

Answers in this form are analyzed anonymously

__________________ participant identifier __________ date ________ revision

9. If you have lived or travelled in countries other than your country of residence or country of origin for three or more months, then indicate the name of the country, your length of stay, the language used,[1] and the frequency of use of the language for each country.

Country	Length of stay[2] *(months)*	Language used	Frequency of use[4] *(circle one)*
			Never **1 2 3 4 5 6 7** Always
			Never **1 2 3 4 5 6 7** Always
			Never **1 2 3 4 5 6 7** Always
			Never **1 2 3 4 5 6 7** Always
			Never **1 2 3 4 5 6 7** Always

1) Other than your mother tongue
2) You may have been to the country on multiple occasions, each for a different length of time. Add all the trips together.
3) If there is not enough space in the table above, continue on the last page
4) Please rate your use of the language according to the following scale (circle the number in the table)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	*2*	*3*	*4*	*5*	*6*	*7*

10. Indicate the age at which you started using each of the languages you have studied or learned in the following environments.

	At home	With friends	At school	At work	Language software	Online games
German						
English						
Swedish						
__________						
__________						
__________						

11. What language do you usually speak with your ... ? *(if more than one language, write all)*

Father __________________________ **Friends** ________________________________

Mother __________________________ **Sisters and brothers** ________________________________

If this family situation does not fit to you, write in your own words.

____

____

Answers in this form are analyzed anonymously

___________________ participant identifier ___________ date ________ revision

12. Indicate the language used by your teachers for instruction at each educational level. *(If the instructional language switched during any educational level, then indicate the "Switched to" language)*

	Language	(Switched to)
Elementary school	____________________	____________________
Middle school	____________________	____________________
High school	____________________	____________________
College / university	____________________	____________________

13. If you have taken any standardized language proficiency tests (e.g., TOEFL), then indicate the name of the test, the language assessed, and the score you received for each. *(If you do not remember the exact score, then indicate an "Approximate score" instead)*

Test	Language	Score	(Approximate score)

14. Rate the strength of your foreign accent for each of the languages you have studied or learned. *(Rate the strength of your accent according to the scale under the table, circle the number in the table)*

Language	Strength of accent
German	None **1 2 3 4 5 6 7** Extreme
English	None **1 2 3 4 5 6 7** Extreme
Swedish	None **1 2 3 4 5 6 7** Extreme
____________________	None **1 2 3 4 5 6 7** Extreme
____________________	None **1 2 3 4 5 6 7** Extreme
____________________	None **1 2 3 4 5 6 7** Extreme

None	*Very weak*	*Weak*	*Moderate*	*Strong*	*Very strong*	*Extreme*
1	*2*	*3*	*4*	*5*	*6*	*7*

15. How much of your day do you spend speaking German, English, and Swedish? *(Give the percentage on average relative to each other. If you use other languages, list them and the percentages below.)*

German ______ %   English ______ %   Swedish ______ %

Other ____

Answers in this form are analyzed anonymously

______________	________	______
participant identifier	date	revision

16. Estimate how many hours per day (on average) you spent engaged in the following activities in each of the languages prior to your arrival in Sweden. *(during the previous year)*

	German	English	Swedish
Watching tv (series/movies)	_______	________	________
Listening to radio/podcasts	_______	________	________
Reading for fun	_______	________	________
Reading for school	_______	________	________
Writing with friends	_______	________	________
Writing for school/work	_______	________	________
Speaking with friends	_______	________	________

17. Estimate how many hours per day (on average) you spend engaged in the following activities in each of the languages after your arrival in Sweden.

	German	English	Swedish
Watching television	_______	________	________
Listening to radio	_______	________	________
Reading for fun	_______	________	________
Reading for school	_______	________	________
Writing with friends	_______	________	________
Writing for school/work	_______	________	________
Speaking with friends	_______	________	________

18. Estimate how many hours (on average) per day you spent speaking with the following groups of people in each of the languages prior to your arrival in Sweden. *(the previous year)*

	German	English	Swedish
Family members	_______	________	________
Friends	_______	________	________
Classmates	_______	________	________
Coworkers	_______	________	________

Answers in this form are analyzed anonymously

__________________ participant identifier   __________ date   ________ revision

19. Estimate how many hours (on average) per day you spend speaking with the following groups of people in each of the languages after your arrival in Sweden.

	German	English	Swedish
Family members	_______	________	________
Friends	_______	________	________
Classmates	_______	________	________
Coworkers	_______	________	________

20. Do you mix words or sentences from different languages when you speak? *(This includes starting a sentence in one language but using a word or a phrase from another language in the middle of the sentence)*

[ ] Yes   [ ] No   *(tick the box next to your answer)*

If you answered "Yes" in the question above, then indicate the languages that you mix with the given people and estimate how often you do so. *(circle the number in the table)*

	Language 1	Language 2	Frequency of mixing			
			Never	Some times	Often	Very often
Family members	____________	____________	1	2	3	4
Friends	____________	____________	1	2	3	4
Classmates	____________	____________	1	2	3	4
Coworkers	____________	____________	1	2	3	4

21. In which language do you communicate best or feel most comfortable in terms of listening, speaking, reading, and writing in each of the following environments? *(try to choose one)*

	Listening	Speaking	Reading	Writing
At home	__________	__________	__________	__________
With friends	__________	__________	__________	__________
At school	__________	__________	__________	__________
At work	__________	__________	__________	__________

Answers in this form are analyzed anonymously

________________ participant identifier ________ date ______ revision

22. How often do you use each of the languages you speak for the following activities? *(Please circle the number in the table according to the scale below)*

Language	Thinking	Talking to yourself	Expressing emotion[2]	Dreaming	Arithmetic[3]	Memorizing numbers[4]
German	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
English	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
Swedish	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

1) Please rate your language use according to the following scale (circle the number in the table above)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	*2*	*3*	*4*	*5*	*6*	*7*

2) This includes shouting, cursing, showing affection, etc.
3) This includes counting, calculating tips, counting for months and days, etc.
4) This includes telephone numbers, ID numbers, etc.

23. What percentage of your friends speak the following languages?

German _____ %

English _____ %

Swedish _____ % *(the percentages do not need to add up)*

24. Please comment below to indicate any additional answers to any of the questions above that you feel better describe your language background or usage. You can also provide any other information about your language background or usage.

____

____

____

____

____

____

Appendix K Study 3: Background questionnaire (Short)

Answers in this form are analyzed anonymously

__________________ participant identifier   __________ date   ________ revision

Should you feel that the answer options for a particular question are not representative of your situation you can include additional details at the end of this questionnaire.

1. Rate your language learning skill. In other words, how good do you feel you are at leaning new languages, relative to your friends or other people you know? *(circle one)*

Very poor **1 2 3 4 5 6 7** Excellent

2. Do you like learning languages? *(circle one)*

Not at all **1 2 3 4 5 6 7** Very much

3. Is it easy for you to communicate with people from other countries? *(circle one)*

Very hard **1 2 3 4 5 6 7** Very easy

4. Rate the strength of your foreign accent for each of the languages you have studied or learned. *(Rate the strength of your accent according to the scale under the table, circle the number in the table)*

Language	Strength of accent
German	None **1 2 3 4 5 6 7** Extreme
English	None **1 2 3 4 5 6 7** Extreme
Swedish	None **1 2 3 4 5 6 7** Extreme
__________________	None **1 2 3 4 5 6 7** Extreme
__________________	None **1 2 3 4 5 6 7** Extreme
__________________	None **1 2 3 4 5 6 7** Extreme

None	*Very weak*	*Weak*	*Moderate*	*Strong*	*Very strong*	*Extreme*
1	*2*	*3*	*4*	*5*	*6*	*7*

5. Estimate how many hours per day (on average) do you spend engaged in the following activities in each of the languages after your arrival in Sweden.

	German	**English**	**Swedish**
Watching television	_______	________	________
Listening to radio	_______	________	________
Reading for fun	_______	________	________
Reading for school	_______	________	________
Writing with friends	_______	________	________
Writing for school/work	_______	________	________
Speaking with friends	_______	________	________

Answers in this form are analyzed anonymously

________________ participant identifier   ________ date   ______ revision

6. Estimate how many hours (on average) per day do you spend speaking with the following groups of people in each of the languages after your arrival in Sweden.

	German	English	Swedish
Family members	_______	________	________
Friends	_______	________	________
Classmates	_______	________	________
Coworkers	_______	________	________

7. How much of your day do you spend speaking German, English, and Swedish? *(Give the percentage on average relative to each other. If you use other languages, list them and the percentages below.)*

German ______ % English ______ % Swedish ______ % Other ________________

8. Do you mix words or sentences from different languages when you speak? *(This includes starting a sentence in one language but using a word or a phrase from another language in the middle of the sentence)*

[ ] Yes [ ] No *(tick the box next to your answer)*

If you answered "Yes" in the question above, then indicate the languages that you mix with the given people and estimate how often you do so. *(circle the number in the table)*

	Language 1	Language 2	Frequency of mixing			
			Never	Some times	Often	Very often
Family members	__________________	__________________	1	2	3	4
Friends	__________________	__________________	1	2	3	4
Classmates	__________________	__________________	1	2	3	4
Coworkers	__________________	__________________	1	2	3	4

9. In which language do you communicate best or feel most comfortable in terms of listening, speaking, reading, and writing in each of the following environments?

	Listening	Speaking	Reading	Writing
At home	______________	______________	______________	______________
With friends	______________	______________	______________	______________
At school	______________	______________	______________	______________
At work	______________	______________	______________	______________

Answers in this form are analyzed anonymously

__________________ participant identifier   __________ date   ________ revision

10. How often do you use each of the languages you speak for the following activities? *(Please circle the number in the table according to the scale below)*

Language	Thinking	Talking to yourself	Expressing emotion[2]	Dreaming	Arithmetic[3]	Memorizing numbers[4]
German	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
English	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
Swedish	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
__________	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
__________	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
__________	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

1) Please rate your language use according to the following scale (circle the number in the table above)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	*2*	*3*	*4*	*5*	*6*	*7*

2) This includes shouting, cursing, showing affection, etc.
3) This includes counting, calculating tips, etc.
4) This includes telephone numbers, ID numbers, etc.

11. What percentage of your friends speak the following languages?

German _____ %
English _____ %
Swedish _____ % *(the percentages do not need to add up)*

12. How much do you agree with the following statements? *(circle your answer)*

a. When I cannot remember a word in Swedish I usually think about a English word.

I strongly disagree **1 2 3 4 5 6 7** **I strongly agree**

b. When I cannot remember a word in Swedish I usually think about a German word.

I strongly disagree **1 2 3 4 5 6 7** **I strongly agree**

13. Please comment below to indicate any additional answers to any of the questions above that you feel better describe your language background or usage.

____

____

____

Appendix L Study 3-4: Stimuli creation process

To create the stimuli for the six conditions, potential word pairs in the target language that share a single translation equivalent in the source language, but not in the nuisance language[61] had to be found. Furthermore, these had to be words that the participants would be expected to be familiar with.

Three frequency lists acted as a starting point for the search: the DeReWo (Perkuhn, Belica, Kupietz, Keibel & Hennig (2009) based on COMAS-II for German, Kelly (Volodina and Johansson Kokkinakis, 2012) based on SweWAC for Swedish, and word frequency list based on COCA (Davies, 2010) for English.

Out of these lists, all but nouns were filtered out. Furthermore, all name-type nouns like place names were filtered out as well as all forms with numerical identifiers, spaces, special digits, or symbols. The German list required some additional work to be comparable with the Swedish and the English lists as it was not manually controlled to the same extent than the Swedish and the English lists. Using a VBA (Visual Basic, a programming language used in Microsoft Excel) script the base forms were extracted from the lemmatized list and the frequencies were calculated based on the sums of the individual forms in the non-lemmatized list. Finally, from the cleaned lists of nouns, the 1,000 most common nouns were extracted for all three languages.

Using Google Translate for all other language pairs with the exception of the German-Swedish for which Google Translate did not seem suited enough and for which Nordsteds Stora Tyska Ordbok was used, six translation lists were created listing all the potential translation equivalents for the 1,000 nouns. In case of more than 20 translation options, the translations past 20 were truncated out. The output was then transformed to a long format including a separate line for each specific translation pair and the lists for the two possible directions for each language pair were combined and duplicates were removed. The filtering for numericals, spaces, special characters, and symbols was repeated.

Those words with multiple translation options were then extracted from the output list for each direction creating six lists of potential pairs. Word pairs that clearly would not fulfil the conditions of limited formal similarity between any combination of the three languages or no single translation equivalent in the nuisance language were manually filtered out. From the remaining list of potential

[61] Nuisance language in this context is the language that is neither the source or the target language in the particular CLI condition.

items, items that were deemed to be most likely to fulfill all the predetermined conditions were extracted and put to further analysis.

In the further analysis, the target words were translated backwards to both the source language as well as the nuisance language controlling that the single translation equivalent exists only in the source language and not in the target language. Items that had a single translation equivalent in both languages as well as ones with clear cognates were ruled out. The resulting items were then piloted by native speakers ($n \approx 10$ for each of the six directions) of either the source or the target language with a high proficiency in the other language and were only accepted if more than half of the piloting participants listed both translation equivalents in a translation task reminiscent of a word association task.

The selected items were then evaluated with respect to the relationship between the target word pair, as well as frequency, and word length. Using these as a matching criteria, baseline pairs were created that should resemble the existing pair with translation ambiguity pair as closely as possible. The potential baseline pairs were also back-translated to both the source and the nuisance language and had to fulfill the criteria of having no single translation equivalent in either language as well as no clear cognate status.

Two filler items were also created so that they should be roughly reminiscent of both the treatment as well as the baseline pairs with respect to relationship status, length, and frequency to the extent that the baseline pairs and the treatment pairs would be as indistinguishable from the fillers as possible.

This resulted, all-in-all, in 360 stimuli item pairs across the six potential directions of influence.

Appendix M Study 3-4: Stimuli

Condition 1

Target: German	Source: English

Condition 2

Target: German	Source: Swedish

Condition 3

Target: English	Source: German

Condition 4

Target: English	Source: Swedish

Condition 5

Target: Swedish	Source: German

Condition 6

Target: Swedish	Source: English

Condition 1

Translation ambiguity

Licht	Lampe	Brief	Buchstabe
Boden	Etage	Rennen	Rasse
Bewerbung	Anwendung	Wechselgeld	Verwandlung
Länge	Dauer	Kopf	Oberhaupt
Gespräch	Vortrag	Stockwerk	Geschichte
Ruhe	Frieden	Spitze	Trinkgeld
Energie	Macht	Tonband	Klebeband
Grund	Vernunft		

Baseline pairs

Strecke	Distanz	Strasse	Kreuzung
Schloß	Grube	Unfall	Weile
Rücktritt	Verwendung	Lippenstift	Trockenheit
Reise	Urlaub	Auge	Turnschuh
Eindruck	Ansicht	Gutschein	Wochenende
Wille	Vorbild	Fahrer	Großkunde
Unwelt	Wetter	Sparbuch	Schulbuch
Spiel	Gedenken		

Filler pairs

Tour	Reise	Treffen	Tagung
Geruch	Geschmack	Woche	Monat
Bogen	Papier	Lehrer	Kerze
Angriff	Wette	Partner	Schlaf
Erziehung	Kompromiss	Kandidatur	Kriterium
Polizeiwang	Häufigkeit	Platzwart	Anhäufung
Klima	Wetter	Strecke	Distanz
Markt	Kampfhund	Kreis	Unterarm
Gedanke	Wunsch	Wunsch	Traum
Bauwerk	Kategorie	Werkstatt	Entwicklung
Nähe	Schuld	Zweck	Umwelt
Anlage	Altarraum	Blut	Beutetier
Änderung	Zweck	Saison	Firma
Oberarzt	Stopfen	Schnauze	Sparbuch
Quelle	Beifall	Stelle	Beispiel

Condition 2

Translation ambiguity

Schatz	Steuer	Zahl	Rede
Gesetz	Mannschaft	Satz	Bedeutung
Zustand	Erlaubnis	Aufgabe	Angabe
Atem	Geist	Zustand	Erlaubnis
Bruch	Verbrechen	Maßnahme	Vorkehrung
Glück	Fahrt	Darstellung	Herstellung
Verantwortung	Haftpflicht	Pflanze	Wuchs
Klage	Mühe		

Baseline pairs

Brunnen	Garten	Buch	Liebe
Reihe	Ausstellung	Auge	Erfahrung
Verlust	Verleib	Hilfe	Zukunft
Zwang	Traum	Tatsache	Sichtweise
Wolke	Herrschaft	Richtung	Turnverein
Sicht	Anfrage	Schauspieler	Verteidigung
Vergangenheit	Seelenheil	Beamter	Olive
Auge	Pfad		

Filler pairs

Gebäude	Verkäufer	Kopf	Trage
Zahl	Hoffnung	Tor	Sicht
Reihe	Mitarbeiter	Gastgeber	Bewußtsein
Spiel	Änderung	Tier	Verfügung
Schuld	Mißtrauen	Wetter	Zuwendung
Antrag	Erfolg	Wissen	Wunsch
Gebet	Wille	Umsatz	Abbruch
Schicksal	Mitteilung	Gedanke	Lebenslauf
Heft	Gegenwart	Scheidung	Hüfte
Zweifel	Schlosspark	Pflicht	Hinterland
Blick	Sinn	Wissen	Gesetz
Krankenhaus	Begeisterung	Hochschule	Überzeugung
Voraussetzung	Keuscheit	Notwendigkeit	Rekordzahl
Graben	Verruf	Theater	Geiz
Narbe	Bund	Herz	Sein

Condition 3

Translation ambiguity

Attendance	Visitor	Attitude	Setting
Memory	Reminder	Court	Dish
Shout	Reputation	Treasure	Darling
Snake	Queue	Castle	Lock
Difficulty	Trouble	Bag	Pocket
Carrier	Strap	Negotiation	Trial
Belongings	Estate	Letter	Writing
Ladder	Head		

Baseline pairs

Measurement	Pitch	Learning	Marriage
Theory	Research	Table	Coat
Beard	Fashion	Happiness	Sunset
Brush	Bristle	Barrel	Cap
Presidium	Committee	Firm	Charity
Shuttle	Rocket	Measurement	Burden
Perennial	Flower	Animal	Heart
Needle	Hand		

Filler pairs

Motivation	Blanket	Consensus	Basket
Surprise	Threat	Meaning	Victim
Charge	Pension	Spring	Adoption
Power	Sand	Money	Iron
Rally	Sentence	Robot	Holiday
Shortage	Treasury	Sunlight	Inmate
Label	Tally	Globe	Vortex
Fabric	String	Blade	Knife
Compression	Pressure	Congregation	Meeting
Bit	Piece	Trip	Visit
Envelope	Stamp	Princess	Crown
Deed	Plot	Lust	Fate
Loss	Deficit	Deal	Promise
Community	Partnership	Business	Engineering
Contradiction	Objection	Transformation	Indication

Condition 4

Translation ambiguity

Jump	Hope	Number	Speech
Ear	Handle	Luck	Tour
Association	Unification	Clock	Bell
Treasure	Tax	Incident	Experience
Noon	Dinner	Wait	Expectation
Ability	Capacity	Sentence	Meaning
Breath	Spirit	Attitude	Front
Roof	Ceiling		

Baseline pairs

Grin	Claim	Place	Step
Fan	Defeat	Right	Path
Confidence	Prevalence	Stair	Pole
Garbage	Food	Opening	Education
Harm	Garden	News	Entertainment
Break	Crime	Surgery	Doctor
Reaction	Partner	Relation	Affair
Sheet	Blanket		

Filler pairs

State	Permission	Image	Performance
Chief	Management	Notice	Allocation
Piece	Contract	Task	Specification
Order	Rank	Village	Gust
State	Permission	Course	Pattern
Weight	Influence	Cut	Part
Defence	Protection	Idea	Shadow
Introduction	Adoption	Prayer	Plea
Expansion	Strain	Direction	Hold
Joint	Line	Representation	Fabrication
Court	Hoof	Law	Squad
Credentials	Grade	Cut	Observation
Capacity	Talent	Wear	Application
Subject	Substance	Apparition	Harbor
Paper	Talk	Suit	Costume

Condition 5

Translation ambiguity

Slott	Lås	Stege	Ledare
Samhälle	Sällskap	Rop	Rykte
Kort	Karta	Budskap	Ambassad
Erkännande	Godkännande	Lag	Besättning
Skatt	Ratt	Minne	Påminnelse
Utgift	Utgåva	Myndighet	Ämbete
Poäng	Prick	Uttryck	Utskrift
Domstol	Maträtt		

Baseline pairs

Flaska	Kork	Trasa	Rörelse
Regering	Diktatur	Val	Motiv
Bil	Maskin	Respekt	General
Motivering	Redovisning	Son	Barnbarn
Musik	Mask	Följd	Avslöjande
Signal	Melodi	Rättighet	Doktrin
Vilja	Impuls	Sjukdom	Trauma
Forskare	Vitlök		

Filler pairs

Problem	Attityd	Företag	Ambition
Not	Höjd	Ton	Ursäkt
Hed	Moral	Hand	Finger
Hus	Värme	Inkomst	Skivbolag
Tillväxt	Immunitet	Version	Redaktör
Produkt	Droppe	Hemlighet	Tillgänglighet
Minskning	Reportage	Dag	Motorväg
Liv	Plånbok	Man	Princessa
Pris	Borg	Spelare	Madrass
Grepp	Proportion	Konst	Entusiasm
Skydd	Självkänsla	Närhet	Empati
Symbol	Beröm	Syssla	Merit
Politiker	Kolumn	Utbildning	Läroplan
Demokrati	Ämbete	Andel	Blomma
Cykel	Hjul	Säsong	Atmosfär

Condition 6

Translation ambiguity

Tillämpning	Ansökan	Golv	Våning
Villkor	Skick	Styrka	Hållfastighet
Kista	Bröst	Nota	Lagförslag
Narkotika	Läkemedel	Ljus	Lampa
Våning	Berättelse	Omsorg	Vård
Rygg	Baksida	Kyckling	Fegis
Omsorg	Uppmärksamhet	Ro	Frid
Uppdrag	Uppgift		

Baseline pairs

Engagemang	Ambition	Kassa	Butik
Händelse	Börda	Analys	Utredande
Gruva	Lager	Åska	Koldioxid
Stycke	Klack	Hylla	Möbel
Arena	Revolution	Underlag	Grund
Varg	Djävul	Anarkist	Galning
Faktor	Omständighet	El	Sinne
Förmåga	Resultat		

Filler pairs

Teori	Teolog	Skada	Terror
Språk	Läxa	Spricka	Tosing
Undervisning	Protest	Guld	Bröllop
Innehåll	Panik	Ökning	Förhöjning
Svamp	Turist	Make	Reportage
Knapp	Jacka	Teater	Gemenskap
Röstning	Val	Gård	Farmor
Grundval	Utgångspunkt	Liv	Sorg
Resurs	Beslut	Kapital	Huvudstad
Beteende	Bibliotek	Tolkning	Instrument
Säck	Påse	Sträcka	Avstånd
Mönster	Modell	Protest	Opposition
Mun	Mynning	Vän	Hemort
Hus	Kolonn	Brev	Bokstav
Fabrik	Tak	Föreskrift	Förföljelse

Appendix N Study 3: Proficiency measures in Swedish

Table N1

Correlation matrix for proficiency measures in Swedish

		TIME	RES	DIAL	CEFR	SAP-L	SAP-S	SAP-R	SAP-W	SAP-C	USE	Y-USE
TIME	*r*	–										
	p	–										
RES	*r*	-0.34	–									
	p	0.08	–									
DIAL	*r*	0.68	-0.23	–								
	p	<.01	0.25	–								
CEFR	*r*	0.68	-0.36	0.56	–							
	p	<.01	0.07	<.01	–							
SAP-L	*r*	0.71	-0.38	0.62	0.71	–						
	p	<.01	0.06	<.01	<.01	–						
SAP-S	*r*	0.60	-0.14	0.56	0.55	0.75	–					
	p	<.01	0.48	<.01	<.01	<.01	–					
SAP-R	*r*	0.49	-0.14	0.55	0.62	0.81	0.78	–				
	p	0.01	0.48	<.01	<.01	<.01	<.01	–				
SAP-W	*r*	0.34	0.10	0.31	0.20	0.47	0.53	0.51	–			
	p	0.09	0.63	0.13	0.33	0.02	<.01	<.01	–			
SAP-C	*r*	0.63	-0.18	0.61	0.63	0.90	0.90	0.92	0.70	–		
	p	<.01	0.37	<.01	<.01	<.01	<.01	<.01	<.01	–		
USE	*r*	0.10	0.17	-0.14	0.04	0.20	0.47	0.34	0.16	0.35	–	
	p	0.63	0.41	0.51	0.85	0.32	0.02	0.09	0.43	0.08	–	
Y-USE	*r*	0.32	0.22	0.40	0.21	0.30	0.38	0.29	0.02	0.30	0.47	–
	p	0.11	0.29	0.04	0.29	0.14	0.06	0.16	0.95	0.14	0.02	–

Note. The matrix above presents both objective as well as subjective measures of language proficiency. TIME refers to progress in the intensive language program, RES to relative achieved learning outcomes on the course, DIAL to the vocabulary placement test from DIALANG, and CEFR to self-evaluations of proficiency based on can-do statements. SAP refers to self-assessed proficiency with L, S, R, W, and C referring to listening, speaking, reading, writing and a compound score of overall proficiency respectively. USE refers to self-estimated percentage of daily use across languages and Y-USE to self-reported years of use prior to the commencement of the intensive course.