

Implicit foreign language learning  
after minimal exposure  
across the lifespan

Dissertation for obtaining the Degree of Doctor at the  
Philosophical Faculty of the University of Fribourg (Switzerland)

Nuria Ristin-Kaufmann

From Baar, Zug (Switzerland)

Department of Languages and Literatures

Section of Multilingualism and Foreign Language Didactics

Approved by the philosophical Faculty on request of professors  
Raphael Berthele (1<sup>st</sup> examiner) und Marianne Gullberg (2<sup>nd</sup> examiner).  
Freiburg, 10<sup>th</sup> June 2014.  
Prof. Marc-Henry Soulet, Dean.

**2014**



Implicit foreign language learning  
after minimal exposure  
across the lifespan

Nuria Ristin-Kaufmann

Cite as:

Ristin-Kaufmann, Nuria (2014). *Implicit foreign language learning after minimal exposure across the lifespan*. PhD thesis. University of Fribourg (Switzerland).



*To my friends and family.*



# Contents

<b>Acknowledgements</b> .....	<b>IV</b>
<b>Abstract</b> .....	<b>VI</b>
<b>List of Figures</b> .....	<b>VIII</b>
<b>List of Tables</b> .....	<b>XI</b>
<b>Abbreviations</b> .....	<b>XIV</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>1.1 How and when do we learn foreign languages?</b> .....	<b>1</b>
<b>1.2 Overall Research Questions and Outline</b> .....	<b>3</b>
<b>2 Theoretical background</b> .....	<b>4</b>
<b>2.1 How do we learn foreign languages?</b> .....	<b>4</b>
2.1.1 Role of input .....	4
2.1.2 Speech perception and segmentation.....	10
2.1.3 Artificial and statistical language learning .....	15
2.1.4 Tutored versus untutored foreign language learning at first exposure ..	20
<b>2.2 When do we learn foreign languages?</b> .....	<b>26</b>
2.2.1 Age and cognitive variables .....	34
2.2.2 Production versus comprehension .....	42
2.2.3 Predecessor studies .....	44
2.2.4 The current studies .....	46
<b>3 A lifespan perspective (Study 1)</b> .....	<b>48</b>
<b>3.1 Introduction and Background</b> .....	<b>48</b>
3.1.1 From language input to learning.....	48
3.1.2 Age and Cognition.....	49
3.1.3 Research questions .....	50
<b>3.2 Methods</b> .....	<b>51</b>
3.2.1 Participants .....	51
3.2.2 Experimental Materials .....	54
3.2.3 Cognitive Tests: additional Materials.....	57
3.2.4 Procedure .....	60

<b>3.3</b>	<b>Data Analysis .....</b>	<b>65</b>
3.3.1	Data cleaning.....	65
3.3.2	Statistical analyses .....	66
3.3.3	Hypotheses .....	68
<b>3.4</b>	<b>Results .....</b>	<b>69</b>
3.4.1	Mean accuracy and reaction times per syllable condition.....	69
3.4.2	Comparison of syllable conditions.....	70
3.4.3	Comparison of performance to chance level.....	72
3.4.4	Words compared to non-words .....	72
3.4.5	Development across the lifespan.....	73
3.4.6	Correlations with cognitive variables.....	75
3.4.7	Summary .....	79
<b>3.5</b>	<b>Discussion.....</b>	<b>80</b>
3.5.1	‘Learning’ after minimal exposure.....	80
3.5.2	Development across the lifespan.....	83
3.5.3	Conclusion.....	88
<b>4</b>	<b>Influence of the input (Study 2) .....</b>	<b>89</b>
<b>4.1</b>	<b>Introduction and Background .....</b>	<b>89</b>
4.1.1	From input to consolidation .....	89
4.1.2	Research questions .....	91
<b>4.2</b>	<b>Methods.....</b>	<b>92</b>
4.2.1	Participants .....	92
4.2.2	Experimental Materials .....	93
4.2.3	Cognitive Test: additional Material .....	94
4.2.4	Procedure.....	94
<b>4.3</b>	<b>Data Analysis .....</b>	<b>95</b>
4.3.1	Data cleaning.....	95
4.3.2	Statistical analyses .....	96
4.3.3	Hypothesis.....	97
<b>4.4</b>	<b>Results .....</b>	<b>98</b>
4.4.1	Mean Accuracy and Reaction times per syllable condition – T1 & T2	98
4.4.2	Across groups – Comparison of syllable conditions (T1).....	99
4.4.3	Comparison to chance level across groups at T1 .....	101
4.4.4	Across groups – Words compared to non-words (T1) .....	101
4.4.5	Between groups – Comparison of non-word conditions (T1).....	105
4.4.6	T1 versus T2 – adults’ versus children’s performance on non-words	110

4.4.7	Within subjects comparisons (T1 versus T2)	112
4.4.8	Correlations with cognitive variables	115
4.4.9	Summary	119
<b>4.5</b>	<b>Discussion</b>	<b>121</b>
4.5.1	Effects of input or of previous intake?	122
4.5.2	Consolidation effects?	123
4.5.3	Cognitive or multilingual advantage?	124
4.5.4	Conclusion	125
<b>5</b>	<b>Summary and General Conclusions</b>	<b>126</b>
<b>5.1</b>	<b>Summary of findings</b>	<b>126</b>
<b>5.2</b>	<b>Theoretical Implications</b>	<b>128</b>
5.2.1	Minimal exposure, implicit and statistical learning	128
5.2.2	Age, L1 influence and cognitive variables	130
5.2.3	Quasi-longitudinal perspective	133
<b>5.3</b>	<b>Comments on Methodology</b>	<b>135</b>
5.3.1	Participant selection	135
5.3.2	LDT paradigm	136
5.3.3	Language background & proficiency	137
<b>5.4</b>	<b>Practical Implications</b>	<b>140</b>
<b>5.5</b>	<b>Outlook</b>	<b>142</b>
<b>5.6</b>	<b>Conclusions</b>	<b>144</b>
	<b>References</b>	<b>146</b>
	<b>Appendices</b>	<b>165</b>
	<b>A - The Mandarin Weather Report</b>	<b>165</b>
	<b>B – The Chinese Lexical Decision Task</b>	<b>172</b>
	<b>C – Statistical Analysis: additional Figures and Tables</b>	<b>174</b>
	Data Analysis Study 1 - Cross-validation	174
	ANOVA Study 1	176
	ANOVA Study 2	176

## Acknowledgements

The creation of this thesis required the help of various individuals and institutions to whom I would like to express my gratitude.

This thesis was written as part of the project “Multilingualism through the lifespan” funded through the Sinergia programme of the Swiss National Science Foundation (SNF-130457, PI Raphael Berthele), whom I thank sincerely, and specifically under sub-project ‘D’ “The effects of first exposure to an unknown language” (Marianne Gullberg).

I also thank Lundbergska IDO-fonden, Sweden, for a grant awarded to Marianne Gullberg, which funded Study 2 in this thesis.

Foremost, I offer my deepest gratitude to my supervisor, Marianne Gullberg, for her continual wise and patient support throughout my thesis. Despite the 1200 kilometers between our workspaces, our weekly virtual meetings were a true enrichment, scientifically as well as humanly and a great opportunity for exchanging ideas and concerns. Conferences and meeting times that we spent together were always very enjoyable, eventful and memorable. I could not have wished for a better advisor and mentor for my PhD.

Special thanks also go to Raphael Berthele for initiating and encouraging the Sinergia project and for providing moral support and encouragement from the first day onward when I applied for the position with an unstable internet connection while peddaling in the Argentinian pampa.

I am very grateful to my colleagues on the Sinergia project, especially to Irmtraud Kaiser, Lenny Bugayong and Jan Vanhove for their collaboration during data collection and collegial words of advice.

I owe a very important debt to Leah Roberts, Christine Dimroth, Aaju Chen, Peter Indefrey and Marianne Gullberg for letting me use their experimental paradigm and

stimulus material that they have prepared with much effort and great care. I want to thank Aoju Chen particularly for re-recording the stimulus material.

I am deeply grateful to Joost van de Weijer for statistical support and advice. Thank you for your patience and prompt assistance in moments of despair, especially if they occurred a few hours before a presentation or paper deadline.

Study 2 could not have been realized equally swiftly if it was not for the help and organisational talent of my student assistant, Ilias Kaufmann.

Study 1 and Study 2 would not happen to be possible without the willingness and endurance of our participants to all of whom I want to express my gratitude. I specially thank Marina Midic, Eva Bühler, Shoshana Bauer, Marco Utiger and Rahel Moos-Dettwiler to wake the motivation of their students and children for participation in our studies.

Thank you to Stephan Schmid and Adrian Leemann for advice on phonetics.

I would like to also thank Muriel Clausen for proofreading this thesis and Martina Zimmermann for helping me with the printing.

The most special thanks go to my family for supporting me throughout everything.

## **Abstract**

This thesis examined how people at different ages approach a new language input, looking specifically at the ability to implicitly acquire phonotactic information after minimal, first exposure to continuous audiovisual speech.

In the second-language-acquisition literature, age as a constraint is hotly debated, while the artificial grammar literature has shown that children as well as adults are able to segment and generalize unknown patterns after only brief exposure. Our goal was therefore three-fold: to cover a bigger age-range (instead of comparing 2-3 different age groups), to examine how natural (instead of artificial) complex speech is processed at first exposure, and to examine the potential influence of cognitive variables and previous language skills operationalised as the number of foreign languages (L2s) known.

In Study 1, we tested whether 152 Swiss-German speaking multilinguals between the ages of 10 and 86 could detect violations to syllable structure in a Lexical Decision Task after listening to a seven minute Weather Report in Mandarin Chinese, and whether they could apply phonotactic knowledge derived from the input to new items of the language. Phonotactically violated three- and two-consonant-clusters and CVC syllables ending in an illegal plosive were significantly easier to reject than CVC syllables ending in an illegal nasal. Overall, participants rejected all consonant clusters (except CV\_nasal) correctly significantly above chance level. Importantly, this ability improved with increasing age. Moreover, crystallized intelligence and number of L2s positively correlated with age and positively predicted higher accuracy in the Lexical Decision Task.

In Study 2, we tested what effect input had on an adult (ages 30-40) and a child (ages 10-11) group compared to two matching control groups that did not receive any input. Adults in the exposure condition performed significantly better than control adults on the critical stimulus condition CV\_nasal, but not on any other syllables. Children did not differ. Both adults and children maintained their level of performance after one week of consolidation, suggesting that adults were able to learn implicitly and that implicit learning effects were embedded in memory.

These results provide evidence for a life-long ability to learn abstract linguistic patterns not only from artificial but also from natural continuous speech already after seven minutes of contact with an unknown language. Additionally, this ability seems to improve with increasing age, which speaks against a simplistic age effect for the perception and generalization of newly acquired phonotactic knowledge to non-native language input, and challenges claims against an adult capacity for implicit learning.

## List of Figures

Figure 1: A video-frame from the Weather Report .....	54
Figure 2: left panel: raw reaction times across all subjects, across all conditions after removal of responses <100ms and >2000ms; right panel: log-transformed reaction times.....	66
Figure 3: Boxplots comparing (left) the mean transformed hit rate and (right) the mean transformed reaction times, averaged across all participants in each syllable condition. Arcsine-square-root-transformed hit rate at the chance level equals 0.79, $n = 152$ .....	70
Figure 4: Boxplots comparing (left) the mean transformed hit rate and (right) the transformed mean reaction times, averaged across all participants in words and non-words. Arcsine-square-root-transformed hit rate at the chance level equals 0.79, $n = 152$ .....	73
Figure 5: Linear models for transformed hit rate in the four syllable conditions and age. Transformed hit rate at the chance level equals 0.79, $n = 152$ .....	73
Figure 6: Linear models for mean transformed reaction times in the four syllable conditions and age, $n = 152$ .....	74
Figure 7: Linear correlation between age and cognitive variables. Top panel from left to right: crystallized intelligence (Gc), fluid intelligence (Gf); bottom panel from left to right: working memory (WM), number of foreign languages without German (#L2), $n = 152$ .....	75
Figure 8: Linear correlation between crystallized intelligence and hit rate for each experimental condition, $n = 152$ .....	76
Figure 9: linear model for fluid intelligence and probability of correct rejection in each experimental condition, $n = 152$ .....	77
Figure 10: Linear correlation between working memory and hit rate for each experimental condition, $n = 152$ .....	77
Figure 11: Linear correlation between number of foreign languages and hit rate for each experimental condition, $n = 152$ .....	78
Figure 12: left panel: raw reaction times across all subjects, across all conditions after removal of responses <100ms and >2000ms; right panel: log-transformed reaction times.....	95

Figure 13: Boxplots comparing (left) mean transformed hit rates and (right) mean transformed reaction times, averaged across all four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1,  $n = 68$ ), for all four conditions (VCCC, VCC, CV\_nasal, CV\_plosive). .....99

Figure 14: Boxplots to compare mean transformed hit rate for words and non-words, for all four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1;  $n = 68$ ). Transformed chance level equals .79. ....102

Figure 15: Boxplots to compare mean transformed reaction times, for words and non-words, for all four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1;  $n = 68$ ). Transformed chance level equals .79. ....103

Figure 16: Boxplots to compare mean transformed hit rate per condition for each group (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1,  $n = 68$ ). Transformed chance level equals .79. ....105

Figure 17: Boxplots to compare mean transformed reaction times, per condition for each group (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1,  $n = 68$ ). .....107

Figure 18: Boxplots to compare the mean transformed hit rate per condition, comparing the two experimental groups at T1 and T2 (Ex\_Ch = Experimental Children T1, Ex\_Ch\_w = Experimental Children T2, Ex\_Ad = Experimental Adults T1, Ex\_Ad\_w = Experimental Adults T2). Ex\_Ch/Ex\_Ch\_w:  $n = 14$ ; Ex\_Ad/Ex\_Ad\_w:  $n = 18$ . Transformed chance level equals .79. ....110

Figure 19: Boxplots to compare the mean transformed reaction times per condition, comparing the two experimental groups at T1 and T2 (Ex\_Ch = Experimental Children T1, Ex\_Ch\_w = Experimental Children T2, Ex\_Ad = Experimental Adults T1, Ex\_Ad\_w = Experimental Adults T2). Ex\_Ch/Ex\_Ch\_w:  $n = 14$ ; Ex\_Ad/Ex\_Ad\_w:  $n = 18$ . ....111

Figure 20: Individual subject comparisons per condition of mean transformed hit rate T1 versus T2 (Ad = Experimental Adults,  $n = 18$ ). Transformed chance level equals .79. ....112

Figure 21: Individual subject comparisons per condition of mean transformed hit rate T1 versus T2 (Ch = Experimental Children,  $n = 14$ ). Transformed chance level equals .79. ....113

Figure 22: Individual subject comparisons per condition of mean transformed reaction times T1 versus T2 (Ad = Experimental Adults, n = 18). .....	114
Figure 23: Individual subject comparisons per condition of mean transformed reaction times T1 versus T2 (Ch = Experimental Children, n = 14). .....	114
Figure 24: Linear correlation between age and cognitive variables; (left) working memory (WM); (right) number of foreign languages without German (#L2), n = 32. ....	115
Figure 25: Linear correlation between working memory and transformed hit rate for each experimental condition, n = 32. Transformed chance level equals .79... 116	116
Figure 26: Linear correlation between number of foreign languages and transformed hit rate for each experimental condition, n= 32. Transformed chance level equals .79. ....	117
Figure 27: Linear correlation between age and cognitive variables; (left) working memory (WM); (right) number of foreign languages without German (#L2), n = 36. ....	117
Figure 28: Linear correlation for working memory and transformed hit rate for each experimental condition, n = 36. Transformed chance level equals .79. ....	118
Figure 29: Linear correlation for number of foreign languages and transformed hit rates for each experimental condition, n = 36. Transformed chance level equals .79. ....	119
Figure 30: Cross-validation plots for all six syllable conditions, comparing a linear and quadratic model, in terms of the arcsine of the mean hit rate, across the lifespan.....	174
Figure 31: Cross-validation plots for all six syllable conditions, comparing a linear and quadratic model, in terms of the log of mean reaction times to correct answers, across the lifespan.....	175

## List of Tables

Table 1: Participant characteristics in terms of the distribution of gender, age, number of foreign languages (Mean #L2s) and which foreign languages (L2s) were spoken by how many participants in each age group, $n = 152$ .	53
Table 2: Four violation types of the 128 non-words used in the Lexical Decision Task.	55
Table 3: Experimental design of ten tasks	63
Table 4: The four non-word conditions in the LDT. The illegal consonants are underlined.	69
Table 5: Mean and standard deviation (in brackets) of untransformed hit rate and reaction times (in ms) on the Lexical Decision Task per non-word syllable condition and real words.	69
Table 6: Tukey's honestly significant difference criterion for mean transformed hit rate in the comparisons of the four conditions VCCC, VCC, CV_plosive and CV_nasal.	71
Table 7: Tukey's honestly significant difference criterion for mean transformed reaction times in the comparisons of the four conditions VCCC, VCC, CV_plosive and CV_nasal.	71
Table 8: Participant characteristics in terms of the distribution of gender, age, average number of foreign languages (Mean #L2s), and which foreign languages (L2s) were spoken by how many participants in each group, $n = 68$ .	93
Table 9: The four non-word conditions in the LDT. The illegal consonants are underlined.	98
Table 10: Mean and standard deviation (in brackets) of untransformed hit rates on the Lexical Decision Task per non-word syllable condition and real words.	98
Table 11: Mean and standard deviation (in brackets) of untransformed reaction times on the Lexical Decision Task per non-word syllable condition and real words.	98
Table 12: Tukey's honestly significant difference criterion for mean transformed hit rate, between the four conditions (VCCC, VCC, CV_plosive and CV_nasal)	

across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1). ..... 100

Table 13: Tukey's honestly significant difference criterion for mean transformed reaction times in between the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1). ..... 101

Table 14: Tukey's honestly significant difference criterion for mean transformed hit rate in the comparisons of words (W) and non-words (NW), across the four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1). ..... 103

Table 15: Tukey's honestly significant difference criterion for mean transformed reaction times in the comparisons of words (W) and non-words (NW), across the four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1). ..... 104

Table 16: Tukey's honestly significant difference criterion for mean transformed hit rate between the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1). ..... 106

Table 17: Tukey's honestly significant difference criterion for mean transformed reaction times between the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1). ..... 109

Table 18: Analysis of Variance to compare the mean transformed hit rates of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive). ..... 176

Table 19: Analysis of Variance to compare the mean transformed reaction times of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive). 176

Table 20: Omnibus ANOVA to compare the mean transformed hit rates of the different groups for words and non-words (VCCC, VCC, CV\_nasal, CV\_plosive). ..... 176

Table 21: Omnibus ANOVA to compare the mean transformed reaction times of the groups for words and non-words (VCCC, VCC, CV\_nasal, CV\_plosive). .... 176

Table 22: ANOVA to compare the mean transformed hit rates of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1). ..... 177

Table 23: ANOVA to compare the mean transformed reaction times of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive) across all four

groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1).....	177
Table 24: Omnibus ANOVA to compare the mean transformed hit rates of the different groups per syllable conditions (VCCC, VCC, CV_nasal, CV_plosive). .....	177
Table 25: Omnibus ANOVA to compare the mean transformed reaction times of the groups per Syllable condition (VCCC, VCC, CV_nasal, CV_plosive).....	177
Table 26: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given VCCC.....	178
Table 27: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given VCC. ....	178
Table 28: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given CV_nasal.....	178
Table 29: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given CV_plosive. ....	178
Table 30: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given VCCC.....	178
Table 31: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given VCC. ....	179
Table 32: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given CV_nasal.....	179
Table 33: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given CV_plosive. ....	179

## Abbreviations

Ad_Co	Adult Control group
Ad_Ex	Adult Experimental group
Ad_Ex_w	Adult Experimental group after one week
AGL	Artificial Grammar Learning
AIT	Autonomous Induction Theory
ANOVA	Analysis of Variance
AoA	Age of Acquisition
AoAr	Age of Arrival
BAF	Barcelona Age Factor project
CAH	Contrastive Analysis Hypothesis
CEFR	Common European Framework of Reference for Languages
Ch_Co	Child Control group
Ch_Ex	Child Experimental group
Ch_Ex_w	Child Experimental group after one week
Cohen's $d$	Cohen's measure of sample effect size for comparing two sample means
CP	Critical Period
CPH	Critical Period Hypothesis
CVC	Consonant-Vowel-Consonant syllable
CV_nasal	Consonant-Vowel-nasal syllable
CV_plosive	Consonant-Vowel-plosive syllable
$df$	Degrees of Freedom
EEG	Electroencephalography
ERP	Event-Related Potential
$F$	$F$ distribution, Fisher's $F$ ratio
$F(v_1, v_2)$	$F$ with $v_1$ and $v_2$ degrees of freedom
fMRI	functional Magnetic Resonance Imaging
FT/FA	Full Transfer/Full Access model
Gc	Crystallized intelligence
Gf	Fluid intelligence
HSD	Tukey's honestly significant difference criterion
i-learning	inductive learning

L1	First language
L2	Second language
#L2	Number of second languages
LDT	Lexical Decision Task
LMH	Less is More Hypothesis
<i>n</i>	Number of participants
MMN	Mismatch Negativity
NLNC	Native Language Neural Commitment
NLM	Native Language Magnet
NW	Non-word
<i>p</i>	Probability
PAM	Perceptual Assimilation Model
PDH	Proceduralization Deficit Hypothesis
PET	Positron Emission Tomography
<i>r</i>	Estimate of the Pearson product-moment correlation coefficient
<i>SD</i>	Standard deviation
SLA	Second Language Acquisition
SLM	Speech Learning Model
SMMT	Sound-to-meaning mapping task
<i>t</i>	Student's <i>t</i> distribution; a statistical test based on the Student <i>t</i> distribution; the sample value of the <i>t</i> -test statistic
T1	(testing) time 1
T2	(testing) time 2
UCM	Unified Competition Model
VCC	Vowel-two-Consonant-cluster syllable
VCCC	Vowel-three-Consonant-cluster syllable
W	Word
WM	Working Memory
WR	Weather Report
WRT	Word Recognition Task



# 1 Introduction

## 1.1 How and when do we learn foreign languages?

It remains a hotly debated topic what adult learners are or are not capable of in language learning, and especially what they can do with input. A fundamental question is how adults break into a foreign language system at first contact, when they have no pre-existing knowledge to draw on, and what they can learn. Perdue (1996, p. 138) noted, “Far too little empirical attention has been paid to the very beginnings of the acquisition process”. Since then, first-exposure studies have focused on two main areas of research: First, the learner’s ‘problem of analysis’, as noted by, for example, Klein (1986). The second area has focused around VanPatten’s (1996) framework of input processing principles. This thesis situates itself within the first strand of research. Many authors (e.g. Christiansen, Allen, & Seidenberg, 1998; Klein, 1986) have noted that the second language learner’s task at first exposure consists of different sub-tasks, such as comprehending the utterance, encoding statistical regularities, and integrating these regularities. The aim of this thesis is to contribute to the question of how the learner tackles these three tasks.

Typically, people argue that children are better language learners than adults. The slogan “Younger is better, older is faster” (Long, 2005) is often justified with the explanation that children’s brains are more flexible and more plastic than those of older learners. However, so far no target language property has been identified that adults cannot acquire (compare e.g. Birdsong, 2005). Rather than the notion “use it, *then* lose it”, a better way to describe the faculty of second language (L2) learning and mastery seems to be “use it, *or* lose it” (compare e.g. Diamond, 1996). In this vein, usage-based approaches stress the importance of frequency, assuming that knowledge of language emerges from actual events of language usage (compare Birdsong & Gertken, 2013). We learn language by forming associations between probabilities of occurrence and form-function mappings (Ellis, 2002, p. 144), and tend to understand new utterances based on how frequent we previously perceived and analyzed such utterances (p. 145). Investigating the very initial stage of L2 perception therefore necessitates the control of learners’ prior knowledge and L2 background in order to ensure that learners are genuine beginners with no experience of the language in question.

The acquisition of an L2 obviously differs from the process of acquiring one's first language (L1). Three main aspects of difference are seen (compare e.g. Ellis, 2002; MacWhinney, 2005). Firstly, in the L1, the conceptual development (knowledge about the world) occurs parallel to the acquisition of language and serves as a starting point in adult second language acquisition (SLA). Secondly, exposure conditions in SLA are typically more formal and less naturalistic (e.g. in school contexts), and more varied (e.g. through interactions with other non-native speakers and teachers, and through more distraction in the acquisition process by various other commitments). Thirdly, there is usually some interference or transfer of the L1 to the L2, which, by converging evidence, is explained by similar brain structures processing the L1 and the L2 (compare e.g. Abutalebi, 2008; Abutalebi & Green, 2007; Friederici, Steinhauer, & Pfeifer, 2002; Reichle & Birdsong, in press; Steinhauer, White, & Drury, 2009). Another central aspect underlying the three already mentioned, is the development of cognitive variables across the lifespan and their various influences on language processes (compare e.g. Baltes, 1987; Bialystok, Craik, Klein, & Viswanathan, 2004; Ellis & Sagarra, 2010; Gathercole & Baddeley, 1993; Park & Payer, 2006; Sparks & Ganschow, 2001). We regard it as important to take the developmental stage of these variables into consideration when examining the L2 perception ability.

While supporters of the so-called Critical Period Hypothesis (CPH) consider these diverging starting points of L1 and L2 learning as unfavorable to SLA, doubters of the CPH regard language learning and its opportunities from a completely different angle. Instead of referring to a monolingual standard of 'nativeness' and the ideal (L1-like) language acquisition process, language development is referenced to individuals and their distinct language repertoires (that are constituted by the number and order of acquired languages). This perspective emphasizes the notion of 'multi-competence' for speakers of more than one language rather than 'deficient' competences in the various L2s (compare e.g. Birdsong & Gertken, 2013; Cook, 1992; Ortega, 2013). In this thesis, multilingualism is regarded as a holistic concept, independent of proficiency and, for reasons of simplification, we quantified this variable as the number of L2s participants indicated to master.

## **1.2 Overall Research Questions and Outline**

The aim of this thesis is to explore the human perceptual learning mechanism across the lifespan by examining what information can be extracted from rich, complex, continuous natural speech, after minimal exposure, by multilingual learners of different ages and different cognitive abilities, without any prior instruction or training. Despite recent contributions to the investigation of input processing at first exposure, there are still gaps to be filled in the understanding of natural L2 perception and acquisition across the lifespan and the influence of cognitive skills. This thesis seeks to fill these gaps by the means of two empirical studies.

Study 1 asked three main questions: First, how quickly can adults distinguish sound regularities in natural language input? Second, if adults can extract abstract phonotactic knowledge, does this ability change across the lifespan? That is, do adults or children learn these things more easily? Third, how does the development of cognition and/or multilingualism help language users grasp the phonotactic structure of an unknown foreign language?

Study 2 asked three additional questions: First, is there evidence of real learning from input or is general inferencing from prior experience or knowledge enough? Second, do children differ from adults in this regard? Third, is there evidence of real learning with consolidation of phonotactic knowledge after one week?

In what follows, we first present the relevant theoretical and empirical background to this thesis. The two empirical studies will then be presented individually, introduced by a brief summary of the relevant literature and research questions, the description of the methods used, the results and the discussion thereof. The last chapter summarizes the studies and provides a general discussion of the findings, together with comments on the methodology, as well as methodological suggestions for future research, and an overall conclusion.

## **2 Theoretical background**

First, we will review theories and studies debating the role of input and the work a learner must perform on it at first exposure, such as perception and segmentation. This line of investigation is related to previous research on the possible difference between ‘input’ and ‘intake’, the role of attention and noticing differences, and the difference between intentional and incidental, explicit and implicit L2 learning. The different research strategies of artificial, classroom and training studies will be outlined.

Second, we will review theories and studies concerned with the notion of age effects in L2 learning, including the Critical Period Hypothesis and studies by supporters and doubters thereof. Likewise, the influence of cognitive variables and multilingualism will be reviewed.

Finally, predecessor and present studies are presented.

### **2.1 How do we learn foreign languages?**

#### **2.1.1 Role of input**

How L2 learning proceeds from the very initial stages of contact with a new language to different levels of proficiency is still a matter of debate. Discussions typically concern adult L2 learners who have already spent plenty of time getting familiar with the L2 and accumulated considerable knowledge. In consequence, there is a scarcity of studies that have looked at the very initial stage of L2 learning (compare e.g. Perdue, 1996) and at what adult learners can or cannot do with input at this stage. The role of input is pivotal because language learning cannot take place without it (Gass, 1997). In an illuminating series of studies, Carroll has discussed how input is processed during L2 learning and how it is incorporated into the learner’s perceptual system (Carroll, 1999, 2000, 2002, 2004). In the Autonomous Induction Theory (AIT), Carroll re-conceptualized second language acquisition (SLA) and proposed an SLA processing framework that involves the acquisition of L2 knowledge representations and L2-appropriate segmentation strategies (2004, p. 234). In this framework, two types of innate and automatically operating mechanisms process input. The integrative processor at the lower level combines

smaller representations into larger units, and the correspondence processor at the higher level moves representations from one level (for example the acoustic level) to the next (for example the phonological level). Rules to categorize and combine representations are activated if this parsing procedure is successful. The two processors form a sequential parsing module process input regardless of its linguistic origin. Therefore, L1 parsing procedures initially also apply to L2 stimuli. Parsing, however, is different from acquisition. Acquisitional mechanisms are triggered when parsing fails as a consequence of missing or inadequate categorization rules. This process is described as inductive learning (i-learning). It triggers the acquisitional mechanisms to revise perceptual and parsing procedures in order to analyse novel stimuli (Carroll, 2002, p. 229). Carroll distinguishes i-learning from inductive reasoning in that i-learning is influenced by symbolic representations from long-term- and working-memory. However, this stage of processing remains largely automatic and outside of conscious control. Attention is rather a result than a prerequisite of preliminary input processing. Although Carroll was cautious to use attention as a blanket term, her notion differs slightly from what Gass (1997) considered to be a required element to convert input into learner internal representations. The contradiction might, however, stem from the differing definitions of intake. The possible difference between ‘input’ and ‘intake’ (e.g. Corder, 1967) are therefore related to this line of investigation.

### ***Input and intake***

Corder (1967) discussed the notion of intake in his seminal paper on how input is perceived, stating:

The simple fact of presenting a certain linguistic form to a learner in the classroom does not necessarily qualify it for the status of input, for the reason that input is ‘what is going in’ not what is ‘available’ for going in, and we may reasonably suppose that it is the learner who controls this input, or more properly his intake. This may well be determined by the characteristics of his language acquisition mechanism. (p. 165).

Input is therefore defined as what is available to be learnt, while intake is what learners cognitively register for further processing. In consequence, the role of input and intake is further related to research on attention and noticing differences (e.g. Ellis & Sagarra, 2010).

### ***Noticing and attention***

The noticing hypothesis proposed by Schmidt (1990) goes back one more step by stating that learners cannot register anything without first noticing it. Schmidt argued

that the subjective experience of ‘noticing’ or a sense of awareness of aspects of the ‘surface structure’ of input is necessary for learning to take place, even if ‘unconscious’ learning in the sense of learning without intention or learning without metalinguistic understanding is possible. Schmidt equated awareness with attention. Viewed in that light, all learning is conscious, because input only becomes intake for learning if it is noticed (for a review on noticing see Truscott, 1998) (for a review on attention see Robinson, Mackey, Gass, & Schmidt, 2012). A study by Ellis and Sagarra (2010) examined ways to overcome learner’s attentional biases during initial L2 learning by manipulating their current attention. They found that adults’ L1 experience with finding relevant cues in the input blocked their attention during initial L2 learning. The authors viewed this phenomenon as jointly responsible for differing levels of success in associative learning between child L1 and adult L2 learners. They proposed that attentional difficulties in L2 learners can be overcome by pedagogical interventions and pre-training of relevant cues. At this point, it is again crucial to examine the very initial state of learning and perception in order to know what influences and moderates noticing and attentional processes at the very beginning. Influences of different learning modes, such as the difference between intentional and incidental and implicit and explicit L2 learning (e.g. DeKeyser, 2003; Hulstijn, 2003; Williams, 2009), are therefore also important to consider while studying L2 perception and segmentation.

### ***Incidental and intentional learning***

Incidental learning means learning without the intention to learn, but by ‘picking up’ while engaging in a variety of communicative activities (for example reading and listening), where the focus lies on meaning. This kind of learning starts from the very first exposure to a new language and is opposed to intentional learning, which is goal-directed and with the focus on the form of the language where deliberate commitment to memory comes to play, for example while trying to remember something (cf. Hulstijn, 2003, p. 349). Incidental and intentional learning were originally tested in stimulus-response-learning experiments where the forming of associations was examined under different instruction settings (e.g. Gagné, 1965). The critical feature was whether or not participants were told that they were going to be tested (Hulstijn, 2003). Hulstijn differentiated two types of designs: a between-group (type 1) design and a within-group (type 2) design. The type 1 design was developed earlier and mostly served to demonstrate that incidental learning exists. There are no instructions to learn and no information that learning is going to be tested. The type 2 design instructed participants to learn some stimuli, but not others

that were then tested unexpectedly, in such a way that participants could be used as their own controls for the difference between incidental and intentional learning. The criteria that participants are not told whether there will be testing or not also applies to implicit learning experiments (see below Williams, 2009).

### ***Implicit and explicit learning***

Implicit learning means incidental learning of complex information without the awareness of what was learnt. In one of the earliest studies of implicit learning of artificial grammar by Reber (1967), implicit learning was described as a process of acquiring knowledge without the intention and the awareness to do so. Reber interpreted implicit learning as an inductive process similar to perceptual learning described in Gibson and Gibson (1955), and as an intrinsic part of language learning and pattern perception (Reber, 1967, p. 863). Again, this process is thought to be involved at the earliest stages of language acquisition.

Krashen (1981) initiated the discussion of the contrast between implicit and explicit second language acquisition and learning with the Acquisition-Learning Hypothesis, claiming a strict separation of acquisition and learning. While he considered acquisition to be a purely subconscious process upon which the improvement in language ability was dependent on, he believed the conscious process of learning to be independent thereof. The most important distinction from explicit learning according to Ellis (1994) is the absence or presence of ‘conscious operations’ (cf. p. 1). Conscious intentions to find regularities in the input that lead to the exploration of underlying concepts and rules are characteristic of explicit learning (Hulstijn, 2005).

The definition of the concept of implicit learning is still developing and subject of controversy. Criteria about the learnt information’s status, content and manner were proposed (cf. Seger, 1994, p. 164): First, the acquired knowledge is not fully accessible to consciousness. Second, the learnt information is more complex than a simple association or frequency count. Third, learning happens incidentally as a consequence of the type and amount of cognitive processing on the stimuli. Moreover, implicit learning is considered to be robust over time and preserved in cases of amnesia (cf. *ibid.*). While the neural basis of explicit learning is hippocampal, (non-hippocampal) structures of the basal ganglia, the association cortex, and of the frontal cortex seem to be involved in implicit learning (cf. p. 184). Using fMRI, Seger, Prabhakaran, Poldrack, and Gabrieli (2000) found that the neural substrates of implicit and explicit learning of an artificial grammar differed. Reber, Allen and Reber (1999) supported the claim that implicit learning and

memory depend on disparate brain areas (cf. p. 482). As opposed to cognitive explicit learning, implicit learning is therefore often described as (incidental) associative learning (Kaufman et al., 2010, p. 323).

Implicit learning research can be categorized in three areas (2003): artificial grammars, sequence learning and the control of complex systems. Within these areas, implicit learning has been tested in many different tasks, for example in artificial grammar learning, visuospatial concept learning, co-variation learning and serial reaction time learning tasks (cf. Seger, 1994). Reber (1967) was one of the first to study implicit sequence learning with an artificial grammar learning task. In two experiments, he showed that participants became increasingly sensitive to the grammatical structure by perceptual learning and that the information could be extended to new stimuli in a recognition task. The grammatical rules that are implemented in artificial grammar learning (AGL; see below) studies are typically very complex. Usually, the amount of learning is not very large, yet, in a very short time, participants perform significantly above chance level (50 percent represents mere chance) – a typical score lies between 55-70 percent (cf. DeKeyser, 2003, p. 319). The acquired quasi-abstract knowledge is unconscious if participants believe that they are guessing (for example in a Lexical Decision Task), but indeed succeed to discriminate new stimuli above chance level. This is called the ‘Guessing Criterion’ (cf. Cleeremans & Dienes, 2008). Implicit learning studies are thus interested in producing quasi-abstract knowledge structures which in turn allow generalizations of the implicitly acquired knowledge to new stimuli (cf. e.g. Cleeremans & Dienes, 2008).

### ***Implicit and incidental learning***

Implicit learning happens always incidentally, but incidental learning does not entail implicit learning (2003). During implicit learning, the perception of information automatically and autonomously triggers the formation of implicit knowledge (cf. Hulstijn, 2007, p. 706). Listening, for example, is mostly an implicit process, because we cannot influence how quickly we process incoming speech (cf. p. 709). The resulting (implicit) knowledge is difficult to express, it cannot be consciously ‘inspected’ or verbalized. As opposed to this, explicit knowledge is knowledge that we (consciously) know and can access deliberately (cf. Dienes & Perner, 1999). The storing of implicit knowledge is designed to allow rapid, parallel processing (cf. Hulstijn, 2007). This processing mode is distinctive of native speakers’ (linguistic)

behaviour, such as fluent listening and speaking (cf. p. 706). Connectionist models<sup>1</sup> were considered adequate to capture the two most important characteristics of implicit learning insofar as the connectionist concept of knowledge is based on statistical associations in the absence of rules (cf. DeKeyser, 2003, p. 329).

Robinson (2002, 2005) compared implicit artificial grammar learning to incidental natural learning of Samoan and examined the influence of individual differences in language learning aptitude, intelligence and working memory. He found that explicit memory was significantly positively correlated with the performance of the incidentally learned Samoan items, but not with the implicitly learned artificial grammar items. The two learning conditions were unrelated in all the post-test measures. Robinson (2010) concluded that incidental natural language learning and implicit artificial grammar learning are related but are different, separable processes. Learning was not significant in either condition, but there was evidence that sensitivity to frequency in the input ('chunk strength') and individual differences exerted a different influence on the two learning conditions. Robinson thereby specified the interpretation of Reber et al.'s (1991) findings. Reber and colleagues (1991) claimed that implicit learning is independent of intelligence while explicit learning is sensitive to individual differences in intelligence. Robinson (2010) concluded that implicit but not incidental learning is related to intelligence, such as intelligence and implicit learning are correlated significantly negatively.

Williams (2005) pointed out the importance of prior knowledge for implicit learning. The ability to extract lexical information from speech input correlated with the knowledge of languages that encode grammatical gender. Learners were able to use miniature, non-instructed language input as training to generalize the animacy of noun determiners to new items, although they reported not being aware of having learned anything. Williams (2009) endeavoured to develop better models and theories of implicit learning and the establishment of implicit knowledge. According to Williams, a more scientific criterion is needed to guide researchers in planning their experimental procedures, since all the models developed so far contain some elements or procedures that make them less reliable. He criticized that the definition of implicit learning as 'learning without awareness' only allows an operationalisation of implicitness through the assessment of awareness, which is a

---

<sup>1</sup> Cleeremans and Dienes (2008) modelled many aspects of implicit learning in order to illustrate associative learning processes, to explore computational principles in exemplar-based models, and to make the comparisons of models and their predictions possible.

subjective mental state. According to Williams, much more research needs to be done to specify exactly what this powerful associative implicit learning mechanism does and how it contributes to learning.

### **2.1.2 Speech perception and segmentation**

#### ***How do we perceive an unknown language?***

Speech perception is a complex process by which language sounds are noted, taken in and interpreted in terms of their features, which are then processed and stored as mental representations of abstract categories (cf. Hulstijn, 2007, p. 708). Electrophysiological evidence suggests influence of language experience on stimulus processing at very early stages (cf. e.g. Näätänen et al., 1997; Winkler et al., 1999; Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005). Nenonen, Shestakova, Huotilainen and Näätänen (2005) found evidence of a strong L1 influence on L2 contrast detection even in a well-learned L2.

#### ***Effects of the L1 on L2 perception***

The field of L1 effects in L2 acquisition has a longstanding tradition, with much of the work focusing on production. While the Identity Hypothesis (cf. e.g. Jakobovits, 1970; Klein, 1986), for example, asserts that the acquisition of one language has no influence on the acquisition of another, the Contrastive Analysis Hypothesis (CAH) (Lado, 1957) claims that the structure of an earlier acquired language largely determines the acquisition of a second language. Lado assumed that similar structures are assimilated with ease, while dissimilar structures present difficulties in the form of ‘negative transfer’, which can result in a significant indicator of a foreign accent in the L2 (cf. also Wardhaugh, 1970). Krashen (1981) formulated the Monitor Theory that put primary importance on the comprehensibility of the input that language learners are exposed to. Kellerman (1983) defined psychotypology as the learner’s perceived distance of the L1 and the L2. Following on from this, much work has focused on transfer or crosslinguistic influence (Kellerman, 1979, 1983; Sharwood Smith, 1986; Jarvis & Pavlenko, 2008; Odlin, 1989, 2003) across various domains. A more recent transfer theory is the Full Transfer/Full Access (FT/FA) model by Schwartz and Sprouse (1996; cf. also Dekydtspotter, Schwartz, & Sprouse, 2006). This model is concerned with cognitive states in the L2, as well as with the question of what the ‘initial state’ in L2 acquisition is. In this respect, the FT/FA model holds that the early L2 state is full transfer of the L1 and is influenced by prior linguistic experience (Schwartz & Eubank, 1996).

Concerning the effect and extent of L1-influence on L2-speech perception, there are differing views (cf. e.g. Klein, 1986). A prominent model is MacWhinney's Competition Model (1987) followed by the Unified Competition Model (UCM) (2005, 2008), dealing specifically with the role of input and the growth of cue strength (how often a form is used) in transfer processes. According to MacWhinney (2005), L1 and L2 acquisition rely on the same mechanisms. Language comprehension is based on the detection of a set of valid cues whose strength is determined by their reliability and availability (cue validity – how reliable is a form-meaning association). Language acquisition, consequentially, is cue-driven learning, a mechanism that is relevant for lexical, phonological and grammatical forms of language in both the L1 and the L2. Since the L1 is already established and repeatedly co-activated, it is thought to entrench cue-driven learning in the L2. Transfer effects from the L1 can be blocked using meta-cognitive strategies. The L2, on the other hand, is parasitic on the L1 for being clustered in similar brain regions as the L1. High similarities between L1 and L2 forms are thought to be facilitative for the acquisition of similar L2 forms, while competition between differing forms might negatively affect successful acquisition. Especially for similar forms, neuro-cognitive representations and processing of the L1 and the L2 are expected to overlap (cf. also fMRI and ERP evidence by Hernandez, Li, & MacWhinney, 2005; Tokowicz & MacWhinney, 2005). An interesting recent development along these lines is the Convergence Hypothesis (Abutalebi, 2008; e.g. Abutalebi & Green, 2007; cf. also Morgan-Short & Ullman, 2012). It proposes that the L2 relies on the same processing mechanisms as the L1 and that L2 processing 'converges' with L1 processing with increased L2 proficiency. In this regard, the hypothesis is largely consistent with the claims of the Competition Model. There is increasing neurocognitive evidence in support of the hypothesis (cf. e.g. Abutalebi, 2008; Abutalebi & Green, 2007; Friederici et al., 2002; Reichle & Birdsong, in press; Steinhauer et al., 2009).

Looking specifically at phonological effects, two frequently cited models make opposite predictions regarding whether L2 sounds that are more similar or more dissimilar to the L1 sound category are easier to perceive. The speech Learning Model (SLM) by Flege (1993) suggested that new (dissimilar) sounds of the L2 are more easily processed than (similar) sounds that share a phonetic category in the L1 and the L2 (p. 1589). The perceived similarity between L2 and L1 sounds consequently determines the probability of category establishment. According to the SLM, advanced L2 users have two ways of processing L2 sounds: either by

establishing new phonetic categories for some L2 sounds or by linking new L2 sounds with similar L1 sounds, which are then processed in the same phonetic category. The establishment of new phonetic categories would thus only be possible for L2 sounds that an L2 user can perceive as dissimilar to the closest L1 sound.

Best (1995), on the other hand, assumed that similarity of the L2 sound to the listener's L1 facilitates rather than hinders discrimination. Listeners were observed to require more attention discriminating non-native consonants or non-linguistic details that did not belong to their native language frame (Best, McRoberts, & Goodell, 2001). Thus, the authors generally regarded a stable L1 to be supportive for the perception of foreign utterances. The Perceptual Assimilation Model (PAM) by Best (1995, p. 194) predicted three ways of processing an auditory articulation. First, if a novel sound is perceived as a potential phonological unit, it will be assimilated as an acceptable example of a native phoneme category. Second, if the articulation is only similar to two or more phonemes, the sound falls somewhere between native phonemes as an uncategorized consonant or vowel. Third, if the articulation bears no detectable similarity to any native phoneme, the sound is categorized as a 'non-assimilable' non-speech sound. Additionally, six pair-wise assimilation types with assigned discrimination levels are possible. PAM predicts infants' developmental progress to reach from detecting only non-linguistic information in speech, through recognizing how phonetic variants fit into language-specific phonetic classes, to discovering functions of phonological contrasts that help to distinguish native words (Best et al., 2001).

Usage-based approaches postulate input to be the driving force of language acquisition, whose regularities are analyzed by learners with various cognitive (non-linguistic) tools (cf. e.g. Zyzik, 2009, p. 49). For example, the language acquisition mechanism is extremely sensitive to usage frequency at all levels of language processing (cf. Ellis, 2002; Ellis, 2012). The repeated analysis of distributional characteristics of the input results in the acquisition of language rules (e.g. structural regularities) and associative learning of form-meaning mappings. Therefore, statistical learning processes (see below 2.1.3) and linguistic input are at the core of scientific inquiry in usage-based studies. Linguistic input, together with cognitive skills, are seen as the crucial factors in the distinction of the L1-L2 acquisition processes (cf. Slabakova, 2013). In this thesis, we are specifically interested in the L2 processing of phonotactics.

### ***Phonotactics – regularities and constraints***

Phonotactics define what combinations of phonemes, consonant clusters and vowel sequences are legal in a language and phonotactic constraints define illegal combinations. They are both highly language specific and therefore have to be learnt in the acquisition of an unknown language (cf. e.g. Dell, Reed, Adams, & Meyer, 2000, p. 1357). Knowledge of L2 phonotactic regularities and constraints affect both L2 speech perception (cf. e.g. the identification of speech sounds Massaro & Cohen, 1983; and the identification of word boundaries McQueen, 1998; Norris, McQueen, Cutler, & Butterfield, 1997; Pitt, 1998) and production (cf. e.g. Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997).

Studies have shown that L2 learners use phonetic and phonotactic information from the L1 to detect word boundaries in the L2. Flege and Wang (1989), for example, compared the listening performance of three Chinese L1 groups on the identification of an English L2 phonetic contrast. The contrast between /t/ and /d/ exists word initially in Chinese, but not word-finally like in English. Cantonese, Mandarin and Shanghainese present different phonotactic constraints regarding the word-final plosives. Participants performed according to the authors' prediction and showed increased sensitivity after training. L1 phonotactic constraints were interpreted to influence L2 syllable processing. Comparing English and Japanese listeners in six phoneme detection experiments, Cutler (1994) showed that both Japanese and English listeners apply their L1-specific pre-lexical processing patterns to L2 input, even if they are not appropriate in the L2. In a summary paper by Cutler (2001), the process of L2-word recognition again emerged to be highly dependent on and constrained by the learner's L1. L1-experience affected listener's expectations of L2 syntax, L2 semantics, and most importantly, L2 phonotactics. Cutler concluded that L2-learners use their L1 phonotactics to segment speech, and that they do this even though it may lead to inefficient processing of the L2, but that it would be possible to inhibit this misapplication with increasing L2 competence.

Broersma (2005) compared Dutch learners of English with English native speakers and showed that lexical activations in L2-learners were increased compared to lexical activations in L1-speakers. The increased activations lead to less efficient phonetic processing. Yet, sounds could still be recognized just as accurately as native speakers did, even if phonetic processing was not native-like. In two studies with a similar design, Broersma and Cutler (2008, 2011) showed that activation in non-native listeners was again higher and involved phantom word competition that

was not present in native listeners. In a comparison of German users of English L2 and English L1 speakers, Weber and Cutler (2006) found that even highly proficient L2 listeners who had acquired phonotactic probabilities of the L2 were not able to prevent L1 phonotactics to interfere. Finn and Hudson Kam (2008) found further evidence for indirect and direct constraints of the L1 in L2 listening and in the ability to use statistical information to segment L2 words. They reused the same set of synthesized English words and non-words in four different experiments to examine whether adult learners can indeed inhibit L1 knowledge and learn to track transitional probabilities of a new system. Despite increased amounts of exposure and indirect information about how to segment words, prior linguistic knowledge interfered with learners' segmentation abilities.

### ***How do we segment unknown L2 speech?***

The process of identifying phoneme, syllable and word boundaries in spoken language is called speech segmentation. Contrary to the expectations we might obtain from reading written words, boundaries between lexical units in most spoken languages are surprisingly difficult to identify. In natural speech, many consecutive words are uttered without pauses between them. As if this was not difficult enough, the way an utterance is split into words can have effects on its meaning. A frequently quoted example in artificial intelligence technology is the phrase 'How to wreck a nice beach you sing calm incense', which sounds very similar to 'How to recognize speech using common sense' (cf. Lieberman, Faaborg, Daher, & Espinosa, 2005, p. 507). The problem can thus not be adequately solved in isolation but requires contextual references, grammar, and semantics.

Listeners use different strategies to segment speech input into separate words, depending on the rhythmic structure of the language. Languages are classified into two or three isochronic (or rhythmic) categories, by implication into syllable-timed, mora-timed, or stress-timed languages. Pike (1972 [1945]) first expressed the idea in 1945 and was supported by many linguists (cf. e.g. Abercrombie, 1967; Ladefoged, 1967), which has contributed to the general acceptance of the theory. Mandarin Chinese, for example, is commonly considered as an example of syllable-timed languages, while German is considered stress-timed (cf. also Bertrán, 1999). In consequence, L2-listeners typically apply the rhythmic strategy they know from their L1 instead of the one appropriate for the L2 (cf. Cutler, Mehler, Norris, & Segui, 1986; Otake, Hatano, Cutler, & Mehler, 1993), which might be a problem in cases where the L1 and L2 do not share the same rhythmical structure (like Chinese

and German). In these cases, non-native listeners encounter more difficulties in dividing the speech stream into separate words. Cutler and colleagues (1986), for example, investigated and found language-specific segmentation processes in French and English listeners. While native speakers of French segmented their language syllable-by-syllable, English speakers used segmentation strategies based on syllable stress. In their follow-up study (Cutler, Mehler, Norris, & Segui, 1989), the authors also examined French-English and English-French bilinguals. They found limits to bilingualism insofar as the dominant language affected whether participants were able to switch from the segmentation strategy suited for the dominant language to the segmentation strategy of the second language or not. Cutler et al. concluded that only one language is basic for any speaker (cf. p. 230). However, different from the monolingual speakers, the bilinguals 'knew' when not to apply the inefficient segmentation strategy.

Phonotactic regularities and constraints in an L2 are first encountered and acquired through listening to the unknown language (cf. Kittredge & Dell, 2011, p. 2679). As outlined above, identifying word and syllable boundaries in order to segment speech of an unknown language is a challenging task for the L2 learner. Phonotactic competence emerges from using language, because language learning is associative learning (2002). Exposure to accurate and adequate input is therefore essential for new associations and L2 acquisition to emerge. One problem with examining learners' work on the input concerns control of learners' prior experience and knowledge. Artificial and statistical language learning studies have solved this problem by controlling the language input.

### **2.1.3 Artificial and statistical language learning**

Artificial language learning studies have examined the influence of statistical properties of language on the acquisitional learning mechanism using completely controlled L2 input. They typically present short strings of often-repeated syllables and then go on to test whether learners have detected regularities in the input (e.g. Peña, Bonatti, Nespor, & Mehler, 2002; Perruchet & Poulin-Charronnat, 2012; Saffran, Newport, & Aslin, 1996). In this way, transitional probabilities between syllables are the only cues for word segmentation. The Artificial Grammar Learning (AGL) paradigm allows researchers to test whether child and adult learners use this kind of phonotactic and statistical information or not.

### ***The AGL paradigm***

Reber developed the first Artificial Grammar Learning (AGL) paradigm (for a review of theories of AGL see Pothos, 2007; 1967). In the standard procedure, participants first undergo a ‘training phase’ where they are shown a series of letter strings that follow a particular complex rule. In the ‘test phase’, they have to classify new sets of strings into rule-governed ones and others that do not comply with the learnt rule. Many researchers argue that the rules of the AGL paradigm are learned implicitly, because they were never explicitly presented to participants and neither were participants able to verbalize them (cf. e.g. Robinson, 2010). The implicit learning mechanism that is assumed behind AGL is statistical learning (see below). Statistical learning describes the ability to extract similarities and transitional probabilities from input. The statistical learning mechanism seems to be domain-general (both visual and auditory) and species-general (occurs in primates and non-primates – cf. e.g. Fitch & Friederici, 2012). In the linguistic domain, statistical learning processes have been used to explain both phonological and syntactical acquisition and are often examined using the AGL paradigm (see further down). The paradigm has also been used to examine language aptitude and individual differences, as well as to investigate which brain structures are involved in the acquisition of syntax and implicit learning. Neurophysiological evidence supports the claim that artificial language learning mechanisms mirror natural language processing (cf. e.g. Christiansen, Conway, & Onnis, 2012; Fitch & Friederici, 2012; Folia, Uddén, de Vries, Forkstam, & Petersson, 2010; Friederici et al., 2002; Opitz & Friederici, 2004). The section that follows reports studies that used the AGL paradigm to study the human learning mechanism. Thereby, we first report some of the most influential statistical learning studies, and then move on to more recent studies that explore the sequential prediction in the nature of the statistical learning mechanism.

### ***A statistical learning mechanism***

Saffran, Aslin and Newport (1996) conducted one of the earliest studies on statistical learning. They examined eight-month old infants and presented them with nonsense streams of monotonous two-minute speech samples. The speech stream consisted either of four randomly repeated three-syllable ‘pseudowords’ in experiment one, or ‘partwords’ in experiment two. After the exposure, infants were tested on the ‘words’ they heard during the exposure and on new ‘words’ that were generated by combining the syllables differently. They listened significantly longer to new ‘words’. Thereby, the authors could show that infants were able to learn the

statistical regularities of the syllable pairings after this minimal input (cf. also Chambers, Onishi, & Fisher, 2003). This result could also be demonstrated with adult learners, where the authors could provoke enhanced segmentation ability by adding a prosodic cue (vowel lengthening) to the transitional probabilities (Saffran, Newport, & Aslin, 1996). With a direct comparison of children and adults in an incidental learning task, Saffran, Newport, Aslin, Tunick and Barrueco (1997) showed that both performed equally well, even in the situation where the only cues for word segmentation were transitional probabilities, and participants were engaged in a cover task. Onishi, Chambers and Fisher (2002) investigated whether adults could acquire unknown phonotactic regularities from brief listening experience. They constructed lists of CVC words and non-words to train the participants on phonotactic regularities. After a brief distraction task, participants were asked to listen and repeat a list of CVC test items. Legal syllables were repeated more quickly than illegal ones. The authors concluded that listeners became sensitive to novel phonotactics within minutes of exposure. Adults insofar generalized the learnt phonotactics to new syllables as they repeated legal new syllables more quickly than illegal ones.

Statistical learning is particularly well documented in the area of lexical acquisition (Saffran, 2003). Saffran, Johnson, Aslin and Newport (1999) examined whether the same effect of statistical learning could be yielded with non-linguistic auditory sequences and found that both adults and infants succeeded as well as they did on syllable strings. Newport and Aslin (2004) and Creel, Newport and Aslin (2004) further showed that adult learners were capable of such computations when the available statistical patterns occur in non-adjacent elements. While Newport and Aslin (2004) used non-adjacent consonants and vowel segments, Creel et al. (2004) used musical tone sequences. Newport and Aslin (2004) suggested that human statistical learning abilities selectively match the constraints that are exhibited by natural languages and are not otherwise limited (cf. also Christiansen, Onnis, & Hockema, 2009; Diehl & Lindblom, 2004; Jakobson, 1969; cf. above Lindblom, 1986). Along similar lines, Creel et al. (2004) inferred that the more general properties of statistical learning constraints appear to apply to both speech and other types of temporally ordered patterns.

### ***Sequential learning***

The AGL paradigm was also used to examine how we learn to make predictions about the sequential structure of language. Peña, Bonatti, Nespor and Mehler (2002),

for example, composed a clever AGL design to distinguish between statistical processes (that are based on frequency and the distribution of elements in a language) and grammatical processes (like using rules). In five versions of their paradigm, they examined the effects of varied amounts of exposure to continuous, meaningless monotonous speech and the insertion of subliminal 25-millisecond gaps within it. They found that participants performed radically different in distinguishing 'rule-words' from 'part-words' when gaps were present in the exposure. What is more, two minutes of exposure yielded almost the same performance as ten, suggesting that generalizations arise very rapidly in the presence of subliminal signals for segmentation. Conversely, greater exposure to unsegmented speech appeared to solidify memory traces rather than generating information about its structure. The authors concluded that two different computational processes were triggered by subtle difference in the signal that provoked two different behaviours, one biased toward the discovery of statistical patterns, and the other oriented toward the discovery of structure. They concluded that the silent gaps made the monotonous speech slightly more similar to natural language. The structure of natural language presents inhomogeneities in the distribution of sounds, words and phrases. Statistical learning may therefore be assimilated to mastering these structures, or seen the other way round, languages only exhibit those structures that learners are able to track (Seidenberg, Macdonald, & Saffran, 2002). According to Saffran (2003), similarities across languages therefore result from constraints on learning and that is why learners compute some statistics more readily than others. Languages that contain predictive dependencies were easier to learn than languages that lack such dependencies (Saffran, 2002). Saffran and Thiessen (2003) concluded that novel regularities which are consistent with the types of patterns found in the world's languages can be learnt successfully, while regularities that are inconsistent with natural language structure cannot.

The 'drive to predict' is seen as a powerful behaviour in learning, providing us with important clues to abstract structures (cf. Elman, 2009, p. 26). Conway, Bauernschmidt, Huang and Pisoni (2010), for example, found word predictability to be the key. They showed that implicit sequence learning could be linked to an individual's ability to predict the final word in English sentences. The better participants became at extracting statistical relationships contained within visual sequences, the better they could predict the final word of auditorily presented sentences that followed the sequential structure that they learnt visually (cf. also Misyak, Christiansen, & Tomblin, 2010). Perruchet and Poulin-Charronnat (2012)

demonstrated that statistical information on its own is sufficient to extract word-like units after only five minutes of exposure to unsegmented artificial language.

As mentioned above, there is increasing neurophysiological evidence showing that natural language processing is mirrored by artificial language learning mechanisms (cf. e.g. Christiansen et al., 2012; Fitch & Friederici, 2012; Folia et al., 2010; Friederici et al., 2002; Opitz & Friederici, 2004). Friederici and colleagues (2002), for example, performed a training study using a small artificial grammar of Brocanto to test the learning and memorization of grammatical rules and words. Thereby, possible transfer-effects of the L1 were excluded and controlled. Training of the artificial grammar elicited L1-like processing strategies in the brain. Tremblay, Shahin, Picton and Ross (2009) provided further evidence that auditory training altered the neural activity during the detection of stimulus-specific cues. Christiansen and colleagues (2012) reported the neural correlates for natural language and sequential learning to be similar, since they found the same ERP component during the processing of a sequential learning task and a sentence reading grammatical judgment task. McNealy, Mazziotta and Dapretto (2006) examined online word segmentation using fMRI. They found significantly different neural activity depending on whether the speech stream contained statistical regularities, statistical regularities and speech cues, or no cues. In a second study, they verified neural patterns in fMRI to the effect that word segmentation had taken place implicitly. They concluded that participants' auditory processing skills were positively correlated with the neural activity indexing the implicit detection of word boundaries.

In sum, artificial language learning studies have investigated statistical, associative and implicit learning mechanisms and the influence of input properties such as frequency, saliency and transparency. These studies have contributed enormously to the understanding of L2 learning mechanisms. Even though neuroscientific evidence has suggested that artificial language processing represents natural language processing, natural languages present us with different challenges at initial stages. In contrast to naturalistic L2 acquisition, most artificial grammars are small, often repeated, frequently trained or simplified for the learners in order to help them break into the new system of rules. Artificial languages are therefore often less complex than natural languages.

#### **2.1.4 Tutored versus untutored foreign language learning at first exposure**

##### ***Classroom studies***

Classroom studies have examined natural language acquisition and the influence of different input properties in naturalistic settings. They examine effects ranging from a few hours of highly controlled input to six years of classroom instruction (e.g. McLaughlin, Osterhout, & Kim, 2004; Muñoz, 2006; Rast, 2008; Rast & Dommergues, 2003; Shoemaker & Rast, 2013). Rast and Dommergues (2003) and Rast (2008), for example, designed stimuli to investigate the effects of word length, word stress, phonemic distance, lexical transparency, frequency and word position. They examined French L1 beginner learners at first exposure of tutored L2 Polish instruction and tested them after a total of eight hours of controlled classroom input that was spread across six weeks. Using a word repetition task, they found that words in sentence-initial or sentence-final positions were more likely to be repeated than those in sentence-medial positions. They also found that sentence position was related to word length insofar as short words that appeared sentence-initially and long words that appeared sentence-finally were easier to repeat than in the reverted relation. These results were replicated in a study by Shoemaker and Rast (2013), together with the absence of a frequency effect after six and a half hours of input. They suggested that first exposure learners were highly dependent on L1 phonological forms and were especially sensitive to the edges of prosodic domains, but that the accuracy of recognition was not specifically based on repetition of lexical items during exposure.

In another classroom study using ERP measures at test, McLaughlin and colleagues (2004) could show that, although participants were unable to discriminate between words and non-words after 14 hours of classroom instruction, recordings of ERPs indicated that the brain made the discrimination. After 63 hours of instruction, ERPs showed that participants' brains even discriminated between semantically related and unrelated target words. Thereby authors showed that the between-subject variability in brain responses was highly systematic. Brain potentials can therefore be used to identify subgroups of learners and to reveal discrete stages of L2 grammatical learning, because they indicate the existence of an intermediate stage in the learning of L2 grammatical knowledge (cf. McLaughlin et al., 2010, p. 124).

Muñoz (2006) compared younger to older students during six years of learning an L2 in classroom. Within the frame of the Barcelona Age Factor (BAF) project, she

analysed the different acquisition outcomes of children starting English L2 at age eight versus children starting at age eleven, receiving equal amounts of instruction. Longitudinal and cross-sectional testing of language skills on ten different tests revealed that older learners significantly outperformed the early starters in most domains, especially in cognitively demanding tasks. Additionally, Muñoz found that morphosyntactic learning seemed to improve around the age of 12, independently of the amount of instruction, but coinciding with the cognitive growth associated with puberty. In conclusion, age of learning was relevant for skills that can be acquired implicitly, given that there is sufficient input for implicit learning to take place. Older learners displayed an advantage in the rate of learning of most skills, which was seen as connected to their superior cognitive development. Once the differences in cognitive development disappear with age, no more differences in proficiency are to be expected (cf. Muñoz, 2006, p. 34). Revising the empirical evidence, Muñoz (2008) pointed out the importance of enough intensity and high quantity input to allow implicit learning, as well as the relevance of good quality input by well-trained teachers and age-appropriate materials. She concluded that input measures are indeed significantly correlated with output measures in the long term, but starting age is not, and if young learners would not receive massive exposure, they will not outperform older learners (cf. also Muñoz, 2011).

### ***Training studies***

The difference between classroom and training studies is that training can be done in the laboratory where exposure conditions, and especially the input, can be controlled and manipulated more carefully. There it is possible to examine how fine-grained differences between the types of evidence that the listeners receive affect their L2 perceptions. Such manipulations can help to understand, for instance, the impact of speech with or without reference to word meaning, or perceptions of minimal pairs versus statistical regularities, as well as the influence of prior linguistic knowledge. Hayes-Harb (2007), for example, compared the effect of two different trainings, one based on statistical information alone, and the other adding the availability of minimal pairs. A third group received no training. The first training session exploited the statistical tendencies in speech with respect to phonemic distinctions and their distribution along an acoustic continuum. The second training raised awareness to a phonemic contrast by presenting two different sound strings that differed only in the novel contrast, but have different meanings that are illustrated with a picture. In a sound discrimination test, she found perceptual learning after the

statistical information only training, but found more accurate perception of novel contrasts if minimal pairs were available during training.

At initial stages, the perception of phonological contrasts is known to be influenced by L1 phonotactics (cf. e.g. Best, 1995; Flege, 1993). Showalter and Hayes-Harb (2013) were interested in how speakers of intonation languages, like English, learn lexical tone in languages like Chinese. They examined how the ability to perceive novel phonological contrasts can be supported by associative memory for sounds. They trained two first exposure groups on pictures of novel objects that were presented simultaneously with Chinese nonce sound forms and the written words in Pinyin. In one group, the written words were marked for the novel tone and in the other they were not. Participants trained on the tone marked words outperformed the no tone mark group. The authors concluded that these participants developed some knowledge of the correspondences between auditory tones and tone marks during the word learning training. What exactly lead to the development of the enhanced performance, however, is not entirely possible to determine, because the tone marks could have also lead to increased noticing processes and thereby to more robust memory representations.

Carroll and Widjaja (2013) trained and examined the learning of three Indonesian number-markings, a feature that is quite different in the participants L1 English, at first exposure. The plural in Indonesian can be expressed either through reduplication, numeral + classifier constructions, or through referring to single and multiple objects with bare noun phrases. Learners could acquire and differentiate the meaning of all three constructions and retain them over a two-week period. The study did, however, not show that learners could freely combine the newly learnt phonemes to create syntactic representations, because this relied on learning of sound-form-picture associations.

Bisson, van Heuven, Conklin and Tunney (2013) investigated whether form-meaning links and subsequent vocabulary learning can also be created through informal exposure to spoken foreign language. They first exposed participants to pictures and words in an incidental learning phase and tested whether this had an impact on the following explicit learning of foreign language translation equivalents. Accuracy and reaction time results from the translation recognition task were compared between five different groups. The ‘multi-session group’, for example, completed the translation recognition task only the next day rather than immediately

after exposure and once again after one week. Thereby, the researchers were able to explore whether the incidentally acquired form-meaning links were transitory or whether they became embedded in memory. Results revealed rapid learning of foreign language words that appeared in the incidental learning phase and, more importantly, the learning effect remained after one day and was even better after one week of consolidation time. Therefore, the learning effect was not transitory and knowledge was integrated over the course of a week, even though knowledge was acquired incidentally.

### ***Consolidation effects in L2 perception learning***

Studies that longitudinally examine effects of L2 perception can verify whether learning effects are only temporary or whether memory traces became embedded. Tamminen and Gaskell (2006), for example, studied the effect of lexical competition and spoken word recognition over a course of eight months. They taught participants novel words at different time points which allowed them, 'en miniature', to examine effects of 'age of acquisition' (AoA). After a familiarization phase of the novel words, participants were asked to complete different tasks that evaluated the learning and consolidation effects from session to session. The tasks consisted of a phoneme-monitoring task, a word repetition task, a lexicalization test (a lexical decision task), a forced-choice recognition test, and on later sessions also a naming test. The authors found robust lexical representations for novel words emerging and remaining as competitors to existing words throughout the course of the study. AoA was difficult to distinguish from frequency. Novel words of early, middle and late AoA were recognized equally fast; no reaction time differences were found. They concluded that AoA and ease of processing are determined by frequency.

Verbal list learning (e.g. Gais, Lucas, & Born, 2006), spatial learning (e.g. Peigneux et al., 2004), and skill acquisition in visual and motor tasks (e.g. Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002) have demonstrated sleep consolidation of memory traces. The role of sleep consolidation is not yet fully understood, but similar retention intervals have yielded different results if sleep was involved. Fenn et al. (2003) used a naturalistic spoken-language learning task and showed that generalizations of phonological categories arose across different acoustic patterns. Training significantly improved recognition performance. However, over the span of a day's retention interval, this performance degraded. After a night of sleep, recognition performance was completely recovered. The authors could thereby

demonstrate that sleep facilitated the recovery and subsequent retention of the learnt material.

Lindsay and Gaskell (2009) examined to what extent sleep played a role in memory consolidation processes. They investigated the lexicalization and memory for novel words by means of a phoneme monitoring task and a stem completion task for exposure and a lexical decision and a familiarity decision task for testing. Each task was done four times by each participant. This allowed the authors to take advantage of known benefits of spaced learning and testing for memory performance. Participants were asked to attend four sessions during the first day and one more fifth session 24 hours after the fourth. Results showed that lexical competition already emerged during the first day, but the magnitude of the competition effect appeared to double in the period containing sleep. No further training was provided on the second day. Therefore, memory consolidation processes and lexical integration do not require sleep, but sleep seems to be sufficient on its own for lexicalization to occur. Given the design with each two rotating exposure and test tasks, it remains to be tested further whether the competition effect of the first day stemmed from enhancements due to spaced learning, spaced testing, or a combination of the two.

Davis and colleagues (2009) also found an overnight consolidation effect, testing lexical competition, repetition, recognition and word meaning rating after two successive days with training of novel words. Using fMRI, the authors explored the neural mechanisms underlying this consolidation effect in a second experiment. They compared neural responses to words that were learned on different days and novel words. Cortical activation was significantly reduced for words that were learned on the previous day, but was similarly high for unfamiliar novel words and for words that were learnt on the day of scanning. This finding was interpreted to be consistent with the hypothesis that phonological representations are modulated by consolidation.

Tamminen, Payne, Stickgold, Wamsley and Gaskell (2010) compared a sleep group with a no-sleep group on an initial test of word learning and again after a week. Both groups spent a retention interval before the initial test, either asleep overnight or awake during the day. The sleep group significantly outperformed the no-sleep group on the initial test, but not on the second test a week later. Novel words had been integrated into the mental lexicon, which was apparent by slower recognition of familiar words in both groups after the first retention interval.

In brief, classroom and training studies showed that learners could indeed use statistical information occurring in natural language. More accurate perception was elicited if novel contrasts were trained or paired with additional information such as meaning and/or tone marks, or if learners displayed higher cognitive development. Thus richer input, higher frequency and cognitive capacity are more important to L2 acquisition than younger age. Findings of systematic brain potentials after a few hours of learning experience further supported this claim. Such systematic learning effects were also demonstrated in consolidation studies, which showed that the lexical integration in memory occurred briefly after training.

## 2.2 When do we learn foreign languages?

### *The Critical Period Hypothesis*

The idea that children are better language learners than adults was originally proposed by Penfield and Roberts (1959). They hypothesized that the L1 could only be acquired within a natural time-window that closes at age nine, within which the child's brain was predisposed for language learning. Age nine was assumed as a turning point after which the ability for adaption and reorganisation would become rigid and fixed. In consequence, the restricted brain plasticity for language learning was also assumed to apply for L2 learning. Lenneberg (1967) considered the restricted time-window as being motivated by the completion of the hemispheric lateralization process and specialization of the left hemisphere for language by the onset of puberty (cf. also Scovel, 1969). Lenneberg labelled the period between two to nine years the 'critical period' (CP) and formulated the Critical Period Hypothesis (CPH), claiming that learners acquire an L2 more successfully before puberty. Furthermore, within the CP, an L2 can be acquired instinctively and without conscious and laboured effort and from mere exposure to the input. Defenders of the CPH believe that less rapid growth of nerve connections and less plasticity after puberty explains the impossibility to reach native-like competence in an L2.

Lamendella (1977) and later Long (1990) proposed a somewhat less radical notion of a 'sensitive period' where pronunciation is supposed to be the skill most influenced by maturational constraints, but the possibility of language acquisition at later stages was not excluded. In a similar vein, Hyltenstam and Abrahamsson (2003) have advanced the idea of a continuous 'maturational period'. They explained the increasingly difficult L2 acquisition with higher age of onset by a generally linear decline in L2 learning potential through maturation. Cases of exceptionally successful L2 learners are attributed to non-maturational factors (cf. p. 574), but the authors are convinced that late learners in principle do not acquire absolute nativelikeness (cf. Abrahamsson & Hyltenstam, 2009, p. 294). The CP-debate circles more around L2 acquisition because a critical period for L1 acquisition is ethically almost impossible to study. Some case studies of feral children nevertheless exist, most famously the one of Genie, who grew up with minimal human contact and therefore displayed delayed L1 acquisition (Fromkin, 1974). Fromkin's case study neither proved nor disproved the CPH for L1

acquisition, however. Yet, we believe that a possible critical period in L1 does not imply the existence of a CP for L2 (cf. Slabakova, 2013).

### ***Critical Period(s)?***

The Critical Period Hypothesis (CPH) has both supporters (e.g. DeKeyser, 2000; Elman, 1993; Johnson & Newport, 1989; Kuhl, 2004; Lenneberg, 1967; Weber-Fox & Neville, 1996) and skeptics (e.g. Carroll & Widjaja, 2013; Dimroth & Haberzettl, 2012; Friederici et al., 2002; Hakuta, Bialystok, & Wiley, 2003; Neufeld, 1977, 1988; Singleton, 2005; Snow & Hoefnagel-Höhle, 1978; Stein et al., 2006). Two characteristics are minimally shared by defenders of the CPH, namely 1) a specified developmental period within which a high level of preparedness for learning of environmental stimuli is present, and 2) a lack of preparedness outside this period (cf. e.g. Colombo, 1982; Hakuta et al., 2003). Therefore, supporters of the CPH must demonstrate the existence of a critical point at which learning outcomes significantly change. However, there is little consensus about what age constitutes this critical point. Kuhl (2004) proposed this point already at nine months. Up to that time, according to her Native Language Neural Commitment (NLNC) Theory, neurons specialize in the sounds of the L1 and synaptic connections for the perception of non-L1 sounds are lost. Moreover, her Native Language Magnet (NLM) Theory explains how learned prototypes of the L1 function like magnets that wrap the perceptual space. As a consequence, the perceptual sensitivity towards non-native speech sounds is reduced and can lead to insensitivity towards acoustic cues that are crucial for the perception of a foreign phonetic contrast. A different processing pattern is therefore assumed for the L2 than for the L1. Krashen (1973) defined age five as the critical point and started to examine differences between younger and older children considering three parameters: ‘route’, ‘rate’ and ‘ultimate attainment’. ‘Route’ relates to the developmental order or sequence in which elements of the L2 are acquired, while ‘rate’ defines the temporal dimension of each step along the route of acquisition. ‘Ultimate attainment’ delineates the end point of acquisition, or full mastery of morphology, phonology and syntax (cf. e.g. Scovel, 1969). Pinker (1994) set the critical point at age six, Lenneberg (1967) and Scovel (1969) at 12 years and Johnson and Newport (1989) at 15 years. However, not many authors offered reasons for proposing the respective critical age point (cf. Hakuta et al., 2003, p. 31).

The CPH is typically challenged either by evidence of native-like competence in the L2 (cf. e.g. Birdsong, 1992; Bongaerts, Planken, & Schils, 1995; Ioup, Boustagui, El

Tigi, & Moselle, 1994), or by behavioural evidence against a qualitative change in learning out-comes around a critical point of development (cf. e.g. Bialystok & Hakuta, 1999; Birdsong & Molis, 2001; Flege, Munro, & MacKay, 1995; Flege, Yeni-Komshian, & Liu, 1999). Snow and Hoefnagel-Höhle (1978) conducted one of the first studies to question the CPH. In a longitudinal study, they tested the natural L2 acquisition of Dutch by English-L1 speakers in five different age categories with nine different test measures at three different times during one year. They found that 8-10- and 12-15-year-olds advanced best in their L2 acquisition after one year of immersion (compared to Dutch native controls), in all the tested skills. The 3-5-year-olds had the lowest scores on all tests (morphology, syntax, vocabulary and comprehension). Adults and 12-15-year olds made the fastest progress during the first few months of learning (cf. Krashen, Long, & Scarcella, 1979). On all measures, except pronunciation, the order of groups, from proficient to poor, was: 12-15, adults, 8-10, 6-7, 3-5. A linear increase with age, but no age effect was found for the pronunciation test at time 2 and 3. On this basis, the authors concluded that their results disproved the existence of a CP for optimal language acquisition between 2-12 years. Two more recent studies (cf. Montrul & Slabakova, 2003; van Boxtel, Bongaerts, & Coppen, 2005) examined advanced L2 learners in their acquisition of L2 grammatical details that are known to be extremely difficult. In both studies, L2 learners had begun to acquire their L2 after age 12, yet both studies reported L2 learners that performed within the range of native control speakers (19 out of 64 participants in Montrul & Slabakova, 2003) and (8 out of 43 in van Boxtel et al., 2005). The researchers concluded that a natively-like command of the L2 system is indeed possible and does not become unattainable after a certain critical age (cf. also Reichle, 2010).

### *Nativeness for phonology?*

The role of age and age of acquisition for the success of L2 learning is a permanent topic of dispute, especially in terms of the acquisition of L2 phonology. Scovel (1969) called the asymmetry between adult ultimate attainment in morphology and syntax and missing native-like phonology the 'Joseph Conrad Phenomenon' in honour of a Polish writer who became famous for books he wrote in his L3 English. Joseph Conrad had acquired English after childhood, wrote fluently, but continued to speak with a strong Polish accent. Therefore, Scovel held that there is no critical period for L2 acquisition of morphology or syntax, but that there is a CP for the acquisition of phonology around age 12. Neufeld (1977, 1978, 1979) designed a series of studies to examine whether it is possible for some late L2 learners to

achieve native-like pronunciation. He subjected L1 English students to an intensive 18-hour laboratory training of the pronunciation of Chinese and Japanese sound patterns. At test, nine out of 20 students were judged to be native speakers of Japanese and eight of Chinese. The studies were, however, criticized by, for example, Long (1990, pp. 266-268) for methodological weaknesses, especially concerning the rating of native-likeness and the instruction to the judges and the fact that no control sample of native speakers was judged on the same stimuli. The validity of the result that adults can produce native-like pronunciation in an L2 was therefore still doubted (cf. also Neufeld, 1988).

The approach of 'global nativelikeness' was first taken and examined by Ioup and colleagues (1994) in an influential case study of two exceptional adult learners of Egyptian Arabic. The two learners acquired the L2 outside the classroom and their non-native background was no longer noticeable by native speakers by the time of testing. The two L2 learners were described as highly motivated, constantly exposed to a naturalistic environment, and with conscious attention on grammatical form. The authors found that one of two examined adults acquired native-like proficiency on various tasks, except for grammatical intuition. A more recent study has further challenged the CPH and the Conrad Phenomenon: Abu-Rabia and Kehat (2004) examined the severity of a foreign accent in 10 L2-Hebrew speaking immigrants and three native controls. Five judges ranked the participants' proficiency and phonology. Length of L2 exposure, type of input, learning styles, self-esteem, motivation, attitude and most importantly the amount of L1 and L2 use accounted for the success in the phonology rating. Some of the participants even reached higher scores than the native speakers.

A series of studies have tried to point out that the quality of L2 input matters a lot for the respective quality of intake and L2 output (cf. e.g. Flege, 2009; Flege et al., 1999; Gullberg, Roberts, Dimroth, Veroude, & Indefrey, 2010; Hyltenstam & Abrahamsson, 2003; Nikolov & Mihaljević Djigunović, 2011; Rast, 2008). Flege and colleagues (1999), for example, remarked that L2 representations and processing will be more native-like if L2 learners receive more (L2) input from native speakers (cf. p. 98). The authors argued that effects of age of acquisition (AoA) indeed disappeared when factors like education and language use (that correlated with AoA effects) were controlled for. Flege (2009) further highlighted the need to consider that L2 input is generally less adequate than input received in the L1. Especially for late learners of an L2, L2 input must be assessed more accurately.

### *Nativeness for phonetics?*

Johnson and Newport (1989) argued that children learn better because of the increased intensity of language learning opportunities that are present in childhood as opposed to adulthood. They examined the effect of different ages of arrival (AoAr) on L2 English learning and ultimate attainment in L2 grammar in Chinese and Korean immigrants to the USA. They set the critical cut-off point at age 17 and classified participants with an AoAr after that point as late arrivals. Among early arrivals, they found a linear decline in performance, but not among late arrivals where they found individual differences instead. Bialystok and Hakuta (1994, p. 69) reanalyzed Johnson and Newport's age effect and found that the linear decline would be significant for both groups if the cut-off age had been set to 20 (cf. also Birdsong & Molis, 2001, p. 241).

Similarly, Coppieters (1987) compared near-natives and natives on grammatical judgements, to examine whether and for which aspects of grammar, competence differences could be found. He concluded that there were significant differences in competence and that none of the L2 speakers were within the native range. However, Birdsong (1992) replicated this study and found evidence of native competence by postpubertal learners. Birdsong and Molis (2001) interpreted Bialystok and Hakuta's findings and their own replication study of Johnson and Newport and Birdsong's replication of Coppieters as evidence against the CPH (cf. also Montrul & Slabakova, 2003; Reichle, 2010; van Boxtel et al., 2005).

However, Abrahamsson and Hyltenstam (2009) have argued that these replication studies have found native-like performance because they were based on language tests that used simple (grammatical) structures and were generally too easy (cf. p. 253). In their opinion, there was no reference of native-like performance in studies that used techniques with enough 'linguistic scrutiny'. In their own study that used very carefully selected, highly advanced L2 speakers according to six elaborated criteria, Abrahamsson and Hyltenstam did not find any L2 speaker in the late age of onset (AO) group (AO 12+) that fulfilled their requirements for 'nativeness' in the detailed linguistic analysis, even though a few of them were perceived as native speakers in the first analysis. Most of the early L2 learners were rated as native speakers in phase one by a panel of native speakers, and only a few in phase two. The authors generally found a strong negative correlation between the perceived nativeness and AO and stated that the average perceived nativeness began around AO 12.

In short, discussions about nativeness in phonology and in phonetics have both stalled at definitions of nativelike competences and the methodological challenges of

controlling environmental factors and the quality of input that affect the acquisition of L2 phonology and phonetics. Both discussions are therefore ongoing and advocates pro or contra CPH are equally represented.

### ***L2 capacity declining with maturation?***

The Less is More Hypothesis (LMH) was originally proposed by Newport (1988; cf. also Newport, 1990). The LMH proposed that increased cognitive capacities, especially increased memory span, are responsible for the observed non-nativelike endstates in late L2 acquisition. According to the LMH, younger children therefore have a computational advantage for L2 acquisition, because cognitive capacities increase gradually during childhood (cf. also ‘executive control’ and ‘hypofrontality’ Thompson-Schill, Ramscar, & Chrysikou, 2009, below in 2.2.1). Dimroth and Haberzettl (2012) countered this hypothesis with their paper called ‘The Older the Better, or More is more: Language Acquisition in Childhood’. In longitudinally collected data on three L1 Russian speaking children learning L2 German after first arriving in Germany, they found that older children needed less time to build up verbal paradigms in the L2 than younger children did in the L1. The authors concluded that, against the predictions of the LMH, increased memory and information processing capacity did not slow down older children. On the contrary, positive transfer appears to have taken place insofar as prior knowledge and faster abstraction abilities as well as advanced cognitive processes seem to have helped older children with L2 acquisition.

### ***Child versus adult L2 acquisition***

Defenders of the CPH have reasoned that L2 acquisition in older L2 learning children is unsuccessful because they no longer rely on innate implicit learning mechanisms (like they do in the acquisition of the L1) but instead use explicit learning strategies (cf. implicit and explicit learning above in 2.1.2 - cf. e.g. Bley-Vroman, 1991; DeKeyser, 2003; Krashen, 1981). DeKeyser (2003), for example, argued that adults are no longer able to learn implicitly after a ‘qualitative shift’ to explicit learning (cf. DeKeyser, 2012, p. 456). Along the same lines, Janacek, Fiser and Nemeth (2012) observed a rapid decrement of implicit abilities around the age of twelve. Paradis (2004) set the decrease of plasticity in the procedural memory for language already at the age of five and declared that the reliance on conscious declarative memory increases from about the age of seven (cf. p. 59). Bley-Vroman (1991) advocated that adult L2 learning is fundamentally different from child L1 development (cf. p. 4). He assumed that L1 and child L2 learners relied on linguistically domain-specific innate universal learning strategies to which adult L2

learners no longer have access and therefore rely on cognitive learning strategies (Bley-Vroman, 1989). Cognitive strategies alone, however, would not suffice in normal adult L2 learners to achieve perfect success (cf. p. 44). Concerning the route of L2 acquisition, therefore, defenders of the CPH hold that children and adults acquire L2s in qualitatively different ways.

Bley-Vroman's (1988) Fundamental Difference Hypothesis was soon countered by Robinson's (1997) Fundamental Similarity Hypothesis, stating that no evidence supports the dissociation between implicit and explicit learning systems in adult L2 acquisition. Indeed, more similarities than differences are found in L2 learners' routes of acquisition (cf. e.g. McLaughlin, 1992; Montrul & Slabakova, 2003; Nikolov & Mihaljević Djigunović, 2006; Snow & Hoefnagel-Höhle, 1978; van Boxtel et al., 2005). It is in the rate of L2 acquisition and ultimate attainment that differences of age of acquisition are put forward. There is increasing evidence in support of Krashen et al.'s (1979) seminal paper stating an initial advantage of older age for the rate, but a disadvantage for ultimate attainment (cf. e.g. Muñoz, 2006; Singleton & Ryan, 2004). Adults have been found to progress faster in the first stages of learning, especially for morphology and syntax, but child starters have been found to outperform adults in the long run and to be able to reach native-like levels of proficiency.

### ***Neurocognitive evidence pro and contra CPH***

Neurocognitive evidence provides both support for the CPH and evidence against it. Weber-Fox and Neville (1996) examined early and late Chinese learners of English with event-related potentials (ERPs) and found that native-like syntactic L2 competence only emerged if the L2 was acquired before one year of age. Similarly, Elman (1993) argued, that greatest learning occurs in childhood, at the time of the most dramatic maturational changes. Elman's training of artificial connectionist networks failed in fully formed 'adultlike' networks, but succeeded in developmentally 'handicapped' networks with limited memory. This was seen as support of the advantage of 'starting small' and the behaviour of the 'handicapped' network was seen as resembling the one of children.

Opposed to this finding, but on a similar line, Friederici et al. (2002) proposed the 'less-is-more' hypothesis. By studying the acquisition of a small artificial language called Brocanto, the authors could show that processing patterns of a foreign language can still yield native-like patterns in adult-ERPs, given that the system of

new grammatical rules to be learnt is small. This would conform to the assumption that language competence in the L2 affects processing patterns more significantly than age of acquisition (e.g. Winkler et al., 1999). In a study comparing naïve Hungarians with Hungarians that are fluent in Finnish as their L2, Winkler et al. (1999) showed that learning an L2 requires the formation of recognition patterns that are specific to the newly acquired L2. The formation of such memory traces is predicted to take place in a time window of 150-200 ms after stimulus presentation. Näätänen and Winkler (1999) consider this sensory memory to be mirrored in the mismatch negativity (MMN): the generation of the MMN is based on the sensory stimulus representation of the deviant, which is automatically compared to the representation of the standard. Fluent, but not naïve Hungarians seemed to have developed cortical memory representations for the Finnish phonological contrasts, as shown in an enhanced MMN. These memory traces enabled them to categorize Hungarian phonemes pre-attentively (Winkler et al., 1999). Such recognition patterns presumably develop gradually with the exposure to the new language (cf. also Näätänen et al., 1997). In a longitudinal ERP-study, Stein and colleagues (2006) found electrophysiological evidence of L2 learning after five months of intense German training. They interpreted their results as evidence for plasticity in the adult L2 acquisition system. Steinhauer and colleagues (2009) found further neurological evidence in favour of L2 proficiency instead of an AoA advantage. At very high levels of L2 proficiency, native-like ERPs were found. Positive effects of the amount of exposure rather than starting age were also demonstrated with primary students' neuroimaging patterns (cf. Ojima, Matsuba-Kurita, Nakamura, Hoshino, & Hagiwara, 2011). Converging evidence suggests that L2 processing patterns assimilate L1 processing patterns with increasing proficiency (cf. also Abutalebi, 2008; Abutalebi & Green, 2007; Friederici et al., 2002; Reichle & Birdsong, in press; Steinhauer et al., 2009) and that the adult brain therefore remains plastic into high age.

### ***Intermediate summary***

The existence of a critical period for the acquisition of L2 language skills continues to be disputed. One aspect of the debate concerns the context where SLA is studied. Typically, age effects in L2 acquisition are reported from immigrant populations that have arrived in a country with another native language at different ages (cf. e.g. Abrahamsson & Hyltenstam, 2009; Coppieters, 1987; Johnson & Newport, 1989; Weber-Fox & Neville, 1996). Most of these studies interpret their results in line with Selinker's (1972) incipient hypothesis that two different language systems are

responsible for L1 and L2 acquisition and that the occasional ('5 %') adult L2 success can be ignored as an exception to the rule. 'Nativelikeness' is another unresolved matter of dispute. Abrahamsson and Hyltenstam (2009) have even argued that it does not occur in late L2 learners (cf. p. 294). However, they set their criteria for nativelikeness so high that even native speakers did not pass the rating of all ten judges to be natively like. Obviously, a monolingual standard of 'nativelikeness' was applied (although, we do not know whether the 20 native speakers who were used as a standard of nativelikeness actually were monolinguals or not). Usage-based approaches to language acquisition reject linguistic normativity and notions of 'nativelikeness' and 'endstates' of L2 learning as being irrelevant to the understanding of L2 use. They assume that knowledge of language emerges from actual events of language usage. Language development is referenced at the level of the individual language rather than being analyzed in terms of conformity to external or idealized points of reference (cf. Birdsong & Gertken, 2013; cf. also the notion of 'multi-competence' hereafter and 'the bilingual turn' Ortega, 2013).

Debates of nativelikeness are also seen as intertwined with the question of context and factors such as input and cognitive maturity and individual differences (cf. e.g. Sparks & Ganschow, 2001). When L2 learning was examined in the laboratory context (e.g. Friederici et al., 2002), for example, or when factors such as levels of education and amount of input were controlled for (e.g. Abu-Rabia & Kehat, 2004; Flege et al., 1999; Snow & Hoefnagel-Höhle, 1978), age effects in L2 acquisition disappeared, or even favoured older learners (e.g. Dimroth & Haberzettl, 2012). Late L2 learners with increased cognitive maturity, for example, were shown to overtake early learners (e.g. Cenoz, 2002; Dimroth & Haberzettl, 2012; Miralpeix, 2007; Muñoz, 2006; Snow & Hoefnagel-Höhle, 1978). This suggests that factors such as experience with the L1 or cognitive processing advantages might actually be favourable rather than detrimental to L2 acquisition. In sum, discussions about age and SLA are still under debate. In a next step, let us consider the development of some cognitive variables that are commonly measured in L2 acquisition studies.

### **2.2.1 Age and cognitive variables**

#### ***Crystallized and fluid intelligence***

Crystallized and fluid intelligence are discrete factors that were originally foreshadowed by Spearman's (1904) theory of general intelligence, or *g*, as an 'eductive' and 'reproductive' mental ability. Based on *g*, Cattell (1963) founded the theory of crystallized and fluid intelligence which are therefore abbreviated as *Gc*

and Gf, respectively, and developed the concept further together with Horn (1967). The claim was that the two factors were independent of each other. However, many authors have noted an interdependence of the two (cf. e.g. Cavanaugh & Blanchard-Fields, 2006).

Crystallized intelligence (Gc) is a measure of information that has been stored in long-term memory, like general knowledge, vocabulary, and other learnt skills (cf. Baltes, 1987; Cattell, 1987; Horn & Cattell, 1967). A steep increase of Gc is expected up until around the age of 20-25 years, followed by a flatter increase up to high age (cf. Baltes, 1987; Cattell, 1987; Horn & Cattell, 1967). Some studies show an earlier decline of Gc, for example at age 65 (cf. Cavanaugh & Blanchard-Fields, 2006). Scores of Gc typically decline earlier in cross-sectional studies than in longitudinal ones, because cohort effects might act as confounds. Longitudinal studies, on the other hand, might be confounded due to prior test experiences (cf. Cavanaugh & Blanchard-Fields, 2006).

Fluid intelligence (Gf) stands for logical thinking and problem-solving capacities, such as inductive and deductive reasoning, and is thought to be independent of acquired knowledge and crystallized intelligence (cf. Baltes, 1987; Cattell, 1987; Horn & Cattell, 1967). On the other hand, Gf is supposed to be closely related to working memory (WM) (cf. e.g. Engle, Tuholski, Laughlin, & Conway, 1999). The authors argue, that Gf, similarly to WM, is supposed to keep representations active in the face of interference and distraction. In general, fluid intelligence is expected to increase until around the age of 20-25 years and then slowly decrease in older age (cf. Baltes, 1987; Cattell, 1987; Horn & Cattell, 1967). Lack of practice as well as age-related changes in the brain are thought to contribute to the decline of fluid intelligence along the lifespan (cf. e.g. Cavanaugh & Blanchard-Fields, 2006; Lee, Lyoo, Kim, Jang, & Lee, 2005). Bugg, Zook, DeLosh, Davalos and Davis (2006) examined subjects from 20 to 89 years and related general slowing down of processing speed and frontal function decline to the decline in fluid intelligence with increasing age.

### ***Working Memory***

Working memory (WM) is a complex construct the details of which are still not fully agreed upon. The term was coined in the 1960s to describe a short-term store or short-term memory, where information is kept active and readily available for the duration of a few seconds, as opposed to long-term memory where information can

be stored for longer periods (Atkinson & Shiffrin, 1968; cf. also Cowan, 2008; cf. e.g. Miller, Galanter, & Pribram, 1960). Additionally to temporal storage, increasing emphasis was put on the notion of manipulation of information necessary for complex cognitive tasks such as language comprehension, learning, and reasoning (Baddeley, 1992).

A popular model is Baddeley and Hitch's (1974) multi-component model featuring a central executive, a phonological loop and a visuo-spatial sketchpad (see below). Others have questioned the theory of WM as a separate module and have argued that short-term memory, long-term memory and working memory only differ from each other in terms of the attentional control dedicated to the different memory representations, but are not separate structures as such (cf. e.g. Cowan, 1995; Oberauer, 2002; cf. also Szmalec, Brysbaert, & Duyck, 2012). In this view, both storage and processing are seen to be engaged in working memory processes (cf. e.g. Cowan, 2008). Inspired by the notion that working memory and long-term memory interact much more closely than initially thought, Baddeley (2000) extended the traditional model by the episodic buffer. Among SLA researchers, language acquisition is considered a prime example of the collaboration between working memory and long-term memory (cf. Szmalec et al., 2012, p. 76), because a newly acquired word form has the same characteristics like a long-term memory trace that was gradually developed from working memory (cf. p. 79; cf. also Page & Norris, 2009). Szmalec et al. (2012) concluded that while L1 and L2 acquisition require the ability to represent serial-order information in working memory, language perception and production rely on attentional control functions (cf. p. 89). One possible definition of working memory is therefore: The capacity for controlled attention in the face of distraction (Engle et al., 1999).

WM capacity is considered to be limited. Miller (1956) first suggested 'the magical number seven' to be the limit of short-term memory capacity. He found young adults' memory span of digits, letters, words or other units to be plus/minus two elements around seven. Most adults are indeed able to repeat about seven digits in correct order. After reconsideration of the difference between repeating words, letters and digits, Cowan (2001) proposed the WM capacity to be around four chunks in young adults (fewer chunks for children or elderly).

Working memory scores are traditionally distributed across the age span in an inverted u-curve, similar to that of fluid intelligence (correlations have been found,

cf. e.g. Engle et al., 1999), meaning an increase of WM for children and a decline of WM for adults (cf. also Borella, Carretti, & De Beni, 2008; Gathercole, Pickering, Ambridge, & Wearing, 2004; Jenkins, Myerson, Hale, & Fry, 1999; Park et al., 2002; Park & Payer, 2006; Salthouse, 1994, 1996). A steeper age-related decline in WM capacity has been proposed for visuo-spatial WM tasks than for verbal WM tasks (e.g. Hale et al., 2011; Park et al., 2002). While Park and Payer (2006) proposed an overall linear WM decline across the lifespan, Hale and colleagues (2011) found no evidence of age-related variance on verbal WM tasks in participants of 20-89 years of age. Alloway and Alloway (2013) confirmed Hale et al.'s finding that working memory skills across the lifespan seem to be driven by domain-specific differences (e.g. verbal versus spatial) and not functional differences. Studying 5-80-year-old participants, they found considerable growth in WM capacity in children and a peak in WM capacity around age 30, and almost no change in WM capacity in adults.

The three components of working memory identified by Baddeley appear to be present in six-year-old children and are thought to increase with age (cf. Gathercole & Baddeley, 1993; Gathercole et al., 2004). Subvocal rehearsal is the factor mainly responsible for a memory span expansion during childhood. The central executive is needed to direct attention towards relevant information and to suppress irrelevant input, as well as coordinating cognitive processes. The central executive sub-component of the working memory system is closely related to the control of attention and thereby executive functions, because the same prefrontal brain areas are involved in these processes (cf. Gathercole, 2008). Finally, the visual sketchpad stores visual and spatial information and is used for constructing and manipulating visual images. This component's development with age appears to be closely related to the development of the other two working memory components, since young children have a greater tendency to remember pictorial information in visual form and older children start to use the phonological loop (cf. Gathercole & Baddeley, 1993; Gathercole et al., 2004). WM and attention seem to be critical in early stages of language acquisition (cf. De Diego-Balaguer & Lopez-Barroso, 2010).

Frontal brain areas have been found to be involved in processes that require the ability to focus and maintain attention in the course of intruding events (cf. e.g. Kane & Engle, 2002). This involves voluntary shifts of attention, a process that is driven 'top-down' by signals from the prefrontal cortex. This brain region's maturational process lasts into the early twenties (Huttenlocher & Dabholkar, 1997). The

Inhibition-Reduction theory (Hasher & Zacks, 1988) suggests that inhibition processes are responsible for age-related increases in younger participants and age-related decreases in older participants. People with greater WM capacity have been found to be better at suppressing such intruding events (otherwise WM capacity would be limited by storing these intruding events; cf. Fukuda & Vogel, 2009). Low attention span is therefore related to low WM capacity and vice-versa.

Not surprisingly, WM is linked to learning outcomes (cf. e.g. Cowan & Alloway, 2008). WM is supposed to be strongly related to the performance on complex cognitive tasks and to measures of the intelligence quotient and has been found to be closely related to fluid intelligence in such a way as to explain individual differences in Gf (cf. Kyllonen & Christal, 1990). Alloway and Alloway (2010) even found WM in 5-year olds to be a better predictor of academic success 6 years later than intelligence.

Tasks to measure WM range from reading comprehension, over reading span, to problem solving tasks. Many linguists use a reading span task developed by Daneman and Carpenter (1980; cf. also Mackey, Philip, Egi, Fujii, & Tatsumi, 2002). The digit span task is widely used and has been part of the intelligence tests originally developed by Binet and Simon (1905) ever since Jacobs (1887) published a series of studies showing that older children could repeat longer strings of digits than younger children. What is more, the digit span task is an aural phonological WM task that goes well with incidental learning during oral interaction (cf. Mackey et al., 2002). Mackey et al. (2002) have found positive correlations of phonological WM and incidental learning.

The backward digit span task is related to verbal short-term memory (cf. e.g. Gathercole, 2008). It involves executive processes and is therefore supposed to be more difficult than the forward task. This should also be mirrored in higher age differences in the backward task (cf. Hasher & Zacks, 1988). Yet, Grégoire and Van der Linden (1997) found no significant effect of age for the difference between forward and backward digit span. Inhibition was also not a crucial contributor to age-related changes of WM across the lifespan in a study by Borella and colleagues (2008). They found a linear decline of WM with age and a quadratic relationship between inhibition and age in 20-86-year-old participants, but inhibition only accounted for a part of the WM decline. Salthouse and Meinz (1995) found a significant reduction in age-related WM decrease after the control of age-related

influences like inhibition (cf. Hasher & Zacks, 1988) and speed-measures (cf. also Salthouse, 1996).

Service (1992) first examined non-word spans in relation to L2 word learning. She found WM to be a significant predictor of L2 proficiency. Another early study that found WM as an important predictor for learning new words was run by Cheung (1996). Participants with higher non-word spans learned L2 words faster. The argument for a positive correlation of higher WM scores and faster L2 learning is plausible, since working memory demands were shown to be higher in L2 processing than in L1 processing (cf. e.g. Golestani et al., 2006; Indefrey, 2006). Kormos and Sáfár (2008) found that WM (measured using backward digit span) correlated with five out of their six measures of L2 ability, including reading, speaking, and listening. In a study by Robinson (2005, 2010), WM scores (measured using reading span) predicted successful incidental learning of Samoan. Martini, Furtner and Sachse (2013) also found relations between WM and incidental sequence learning in a serial reaction time task as well as between WM and a free generation task. Learning was reduced when timing constraints were introduced. In a study by Mackey and Sachs (2012) that examined 65-89-year-old L2 learners, only learners with high WM scores in an L1 listening-span test showed L2 development. Similarly, Alloway (2009) showed that WM (but not intelligence) was a predictor of learning outcome on two standardized learning measures for children with learning difficulties. Higher WM capacity is also associated with higher L2 proficiency in a study by van Hell and Tanner (2012).

### ***Executive Control***

Cognitive executive control is a term that includes working memory, problem solving, planning, and similar regulatory processes that supervise and manage other cognitive processes. These processes are typically associated with prefrontal areas in the brain (see review in Chan, Shum, Touloupoulou, & Chen, 2008). The frontal cortex is the brain region with the slowest maturation process that extends into mid-adolescence (Huttenlocher & Dabholkar, 1997). Therefore, preschool children are not yet equipped with fully mature executive functions and some of their errors are related to these emerging abilities (Espy, 2004). Preadolescence is characterized by certain growth spurts in executive functions (De Luca & Leventer, 2008), increased response inhibition and selective attention ability (Anderson, Anderson, Northan, Jacobs, & Catroppa, 2001). Eventually, the different brain systems become better integrated during adolescence and the implementation of executive functions and

inhibitory control processes improves (Leon-Carrion, García-Orza, & Pérez-Santamaría, 2004). Myelination in the prefrontal cortex and executive functioning is at its peak at age 20 (De Luca & Leventer, 2008). Executive control functions and cognitive flexibility remain stable up until around the age of 70 in normally functioning adults (De Luca & Leventer, 2008).

Language studies have shown a bilingual advantage for executive control processes, especially more flexible inhibitory and task switching processes up to high age (e.g. Bialystok, Craik, et al., 2005; Bialystok et al., 2004; Bialystok, Craik, & Luk, 2008 ; Craik, Bialystok, & Freedman, 2010), that might even delay the onset of Alzheimer disease (Craik et al., 2010). Other studies have seen an advantage for language learning in children precisely because of less executive control or so-called ‘hypofrontality’ (Thompson-Schill et al., 2009; cf. also the Less is More Hypothesis (LMH) Newport, 1988). In consequence, it proves difficult to compare children and adults’ executive control functions and it usually requires tasks with different degrees of difficulty.

### ***Number of L2s***

A frequent assumption in SLA is that knowledge of other L2s will affect and perhaps facilitate the acquisition of additional languages. Along these lines, Cook (1992) formulated the concept of multi-competence, meaning that the knowledge of more than one language in one person’s mind affects also other competences of that person. The number of L2s a person speaks is seen as one connected system within a multilingual individual, rather than separate or aggregated systems or competences. Cook speaks of L2 users (irrespective of their proficiency level) instead of L2 learners to avoid speaking of deficiencies. He assumed three central characteristics in L2 users. First, L2 users’ knowledge of an L2 is different from native speakers’ knowledge of that language. Second, L2 users’ knowledge of their L1 is no longer the same as that of monolingual speakers. Third, L2 users think in different ways than monolinguals.

In a similar vein, Bialystok and Martin (2004) and Bialystok et al. (2005), for example, have shown that bilingual children are more skilled at some aspects of language learning compared to monolingual children, and that highly proficient bi- or trilinguals have enhanced executive functions. Further advantages of bilingualism are seen in a greater faculty for creative thinking and richer cultural experiences (Beardmore, 2008) as well as generally greater language awareness. Yelland

(1993), for example, showed that one L2 lesson per week raised children's awareness for their L1. Sophisticated mechanisms are assumed to prevent cross talk in multilingual brains (cf. e.g. Dehaene, 1999). Evidence suggests that executive control processes might also be involved in verbal processes such as ordering competing morphological and phonological activations in the multilingual brain (Bialystok, 2011). Moreover, greater grey matter density has been shown in multilingual individuals, and it is argued that learning multiple languages increases the brain's plasticity (Hyashizaki, 2004). Research speaking against a bi- or multilingual advantage typically argues that language switching may be responsible for a detrimental increased processing cost (e.g. Gollan & Ferreira, 2009; Gollan, Montoya, & Werner, 2002; Hernandez & Kohnert, 1999), and experiences of more tip-of-the-tongue retrieval failures (Gollan & Silverberg, 2001).

Kavé et al. (2008) examined whether the number of languages a person speaks predicts the performance on two cognitive screening tests. In a 12-year longitudinal study, they interviewed 814 participants of the oldest Israeli Jewish population (mean age 83 years). Participants were classified into groups of bilinguals, trilinguals and multilinguals (speaking four languages and more). They found a significant language-group effect on all three screening waves. Age, gender and education significantly contributed to the prediction of cognitive state on the first two waves, but none of these variables were significant on their own on the third wave. Neither place of birth nor age at immigration contributed to the prediction of cognitive state. But the number of languages was a significant predictor on all three waves and was more influential than age and education. Interestingly, participants who reported speaking a language best that was not their mother tongue, scored better on the cognitive screening than those who reported their mother tongue to be their strongest language.

Perquin and colleagues (2013) also examined elderly (65+-year-olds) participants who all spoke between two to seven languages. They used a retrospective nested case-control design to examine proxies of multilingualism such as the number of languages practiced, age of acquisition and duration of practice. Specifically, temporal patterns of acquisition and the resulting sequential practice of several languages across the lifespan were of interest. The earlier in life participants reported practicing multilingualism, the more effectively they were protected against cognitive impairment (without dementia). Already a one-year delay to reach

multilingualism was seen as a multiplied risk factor. The protection was interpreted in relation to increased brain plasticity during aging.

### **2.2.2 Production versus comprehension**

We have outlined theories about age effects on L2 learning and how cognitive variables develop with age and thereby also effect L2 learning. Since this thesis investigates effects of first exposure on perception, most of the literature discussed has dealt with perceptual learning in L2 acquisition and possible constraints and prospects. A subsequent topic of investigation would obviously be what implication perceptual learning effects have on L2 production. Although there is no absolute consensus about the description of the interaction between perception and production during L2 phonological acquisition, most researchers consider that – overall – perception precedes production (cf. e.g. Escudero, 2005). Different relationships and interactions between comprehension and production are suggested in various models of L2 perception. These are very briefly touched upon. In a last step, we present the studies that have lead to the topic of this thesis and the present studies.

The Motor Theory of speech perception by Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967; Liberman & Mattingly, 1985), for instance, assumed that the auditory perception of sounds was ‘analyzed’ by the same processes (‘the vocal tract gestures’) that would be involved in the production of the respective sound. The theory has, however, been criticized for not being able to explain how acoustic signals are translated into production (e.g. Hayward, 2000). Since the discovery of the mirror neurons that link the production and perception of motor movements, this theory has gained more interest outside the field of speech perception (cf. e.g. Galantucci, Fowler, & Turvey, 2006).

While the Identity Hypothesis (cf. e.g. Jakobovits, 1970; Klein, 1986), the Convergence Hypothesis (e.g. Abutalebi, 2008; Abutalebi & Green, 2007) and the Competition Model (MacWhinney, 1987) assert that L1 and L2 acquisition may conform to similar patterns, the Contrastive Analysis Hypothesis (CAH; Lado, 1957) claims that patterns similar to those of the L1 will be acquired more easily. Both the Speech Learning Model (SLM) by Flege (1993) and the Perceptual Assimilation Model (PAM) by Best (1993) are also L1-biased in that they suggest that L2 production accuracy will be affected by how well sounds can be perceived and therefore how dissimilar or similar L2 phonetic categories are to the L1 (cf. also

Flege, 2003). Flege, Frieda and Nozawa (1997), for example, found that the amount of L1 use had an effect upon the persistence of a foreign accent in an L2 even if the L2 had been learnt in childhood and had been spoken for many years. The authors discriminated between a high and a low frequency L1 speaking group. A more noticeable influence of the L1 was recorded if the L1 was used more frequently than the L2. The strength of the L1 representation at the time of L2 learning influenced L2 production accuracy more strongly than the age of acquisition of the L2. The importance of the quality of L2 input for respective quality L2 output has been stressed by this line of research (cf. e.g. Flege, 2009; Flege et al., 1999; Gullberg et al., 2010; Hyltenstam & Abrahamsson, 2003; Nikolov & Mihaljević Djigunović, 2011; Rast, 2008).

The Proceduralization Deficit Hypothesis (PDH; e.g. Paradis, 2004; Ullman, 2004) is mostly supported by defenders of the Critical Period Hypothesis (CPH), assuming that effective proceduralization in a L2 is not possible and L2 production as a consequence remains forever slow and non-fluent (cf. also Scovel, 1969). Successful proceduralization by L2 learners has, however, been demonstrated with simple and reliable cues (cf. e.g. Friederici et al., 2002; Tokowicz & MacWhinney, 2005).

Connectionist models promote the role of experience as a main motor of emergent processes (cf. e.g. Elman, 1990, 1993). In this research tradition, Dell et al. (2000), for example, have shown that the language production system quickly adapts to phonotactic experience. Evidence suggests that knowledge gained during perceptual learning transfers to the production domain (cf. e.g. Altenberg, 2005a; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Kittredge & Dell, 2011). These findings seem to be approved by neuroscientific findings showing that speech production activates the same set of regions as speech comprehension, with activation in additional regions for specific production tasks (cf. Price, 2010).

Since we presuppose that perception precedes production, we attach all the more importance to investigating the very initial stages of perception of natural language ‘in the wild’, in order to understand L2 learning and its difficulties. Quality of input matters in order for intake to take place and noticing to happen. To study what learners do with L2 input, we need to take into account input properties such as the rhythmic structure of language, phonotactics, saliency, transparency and frequencies, as well as learner’s prior experiences. A few predecessor studies have tried to tackle exactly this problem.

### 2.2.3 Predecessor studies

Classroom and training studies have highlighted the importance of the quality of input and have shown what learners can do with natural language after only few hours of instruction or a few rounds of training. The question remains, however, what learners can do without any assistance when input has been pedagogically prepared or trained, with pre-existing knowledge completely controlled. How quickly can they break into a new language system at first contact and what type of information can learners extract from real, complex, unmodified speech that is repeated just once? A series of studies (Gullberg et al. 2010; Roberts, Dimroth, & Gullberg, 2010; Veroude, Norris, Shumskaya, Gullberg, & Indefrey, 2010; Gullberg, Roberts, & Dimroth, 2012) have tried to tackle this problem. Inspired by an unpublished pilot project (Zwitserslood et al., 1994), they constructed seven minutes of controlled, but natural audio-visual speech in Mandarin Chinese that allowed them to test the role of item frequency, speech-associated gestures, and word length. Frequency was operationalized as the number of tokens of the target word types in the sample (2/8 tokens), gestural highlighting as whether the word was accompanied by a gesture or not, and word length distinguished monosyllabic from disyllabic words. These components were built into the continuous speech that was presented in the form of a Chinese Weather Report (WR). The authors could thus examine the stepwise development of segmental, phonotactic, and lexical knowledge after minimal exposure.

Gullberg et al. (2012) analyzed the performance of Dutch adults on a word recognition (WRT) and a sound-to-meaning mapping task (SMMT) immediately after the exposure to the seven minutes of speech. They found that learners were able to recognize words and identify relevant noun meaning and map it onto forms. Furthermore, Gullberg et al. (2012) entered 'number of L2s' (that participants reported), and 'response type' (yes/no) into the analysis as control variables, but removed them from their model analysis because they did not affect their model in any way. Their findings showed a significant difference between token frequency of the target words, a significant effect for di- versus monosyllabic words, and an effect of word-internal transitional probabilities for frequent disyllabic words in the WRT. No effect of word-external transitional probabilities, no difference between the numbers of exposures and no difference for gestural highlighting was found for the WRT. For the SMMT, a significantly higher effect of accuracy was found for disyllabic and for gesturally highlighted words, and an effect of higher word-internal transitional probabilities.

Roberts et al. (2010) used the same exposure material for their experimental group and compared them to a control group that did not receive any exposure. Moreover, they compared a child and an adult experimental group. Immediately after the exposure, they tested participants on a Lexical Decision Task (LDT) and compared their performance to that of a control group. The experimental groups performed significantly above chance on the critical items while the control group's performance was at chance. All participants, including the control group, correctly rejected the three- and two-consonant cluster syllables that were supposed to be easier to be rejected as non-words. In a second experiment, Roberts and colleagues compared the performance of the adult experimental group that had received one exposure to a new adult group that received double exposure to the WR before performing the LDT. They could show that adults' performance on the critical CVC syllables increased further after 14 minutes of exposure.

An fMRI study was performed by Veroude and colleagues (2010) who also used the same exposure material (WR) and tested participants on the same WRT as in Gullberg et al. (2012). In a first step, participants were asked to close their eyes during a five-minute resting state period. In a second step, they saw the WR for the first time, and the third followed a second resting state period, followed by a second exposure to the WR in a fourth step. The final fifth step consisted of a third resting period. After the fMRI session, participants were asked to perform the WRT and a six alternative forced-choice semantic decision task. Comparing the three resting state periods, the authors found structural neurological adjustments in functional connectivity before and after the double-exposure to the seven minutes of speech. Furthermore, they found that these adjustments were stronger in learners than in non-learners (defined by their score above chance in the WRT – score at chance = non-learner) during the first and the third resting state.

These studies suggest that adults are capable learners even if input is 'naturally' rich (meaning consisting of many types and few tokens) and as brief as seven minutes. What these studies do not show, however, is the development across the lifespan, the effect of reaction time, and correlations with cognitive variables. The authors (Gullberg et al., 2012; Gullberg et al., 2010; Roberts et al., 2010; Veroude et al., 2010) have not found significant differences in reaction times in any syllable condition and have not tested any cognitive variables, nor have they found an effect for the number of L2s.

#### **2.2.4 The current studies**

The studies revised thus far have shown how adults' and children's perceptual systems are capable of segmenting L2 input at the very initial stages, given that they noticed it (cf. Section 2.1.2). Moreover, classroom, training and consolidation studies showed that adults are able to learn artificial and natural L2 rules and sequences implicitly (cf. Section 2.1.4). Both context of L2 acquisition and quality of L2 input are important (cf. Section 2.1.1 & 2.2).

Second language acquisition (SLA) benefits from the development of cognitive variables with increasing age up to early adulthood: attentional and inhibitional abilities increase with the maturation of the prefrontal cortex, along with enhanced executive control, fluid intelligence and working memory capacities, and generally larger mental lexicons and higher crystallized intelligence (cf. Section 2.2.1). These findings are in line with the view that frequent exercise of L2s rather than early age of acquisition is the determining factor in SLA. As soon as we learn an L2 additional to our L1, we cease to be monolingual and instead acquire multiple competences. In consequence, the evaluation of the respective L2s with a monolingual standard is inappropriate.

Influences such as quality of input and cognitive maturation are equally strong on production as on perception, since the latter occurs before the former. This view is encouraged by findings of overlapping activation patterns in the brain during both processes (cf. Section 2.1.2 & 2.2.2).

In sum, previous studies investigated the following three factors: 1) segmentation ability of natural language at the very initial stage of L2 acquisition, 2) the influence of age, and 3) the influence of the development of cognitive variables and knowledge of other languages. If natural language was considered, it was either simplified or previously trained. The age span was either only investigated in a few groups or the experimental design did not permit to investigate the very initial state of L2 acquisition.

In the current studies, we examine the very initial stage of L2 acquisition using complex natural language without any simplifications (factor 1) across almost the whole lifespan (factor 2) and under control of cognitive variables and prior linguistic knowledge (factor 3). To the best of our knowledge, no study considered all three factors in combination.

In Study 1 of this thesis, we cover almost the whole age spectrum in order to analyze any shifts across the lifespan in the ability to extract L2 phonotactic information after minimal exposure. Moreover, we compare the performance on cognitive variables across the lifespan and examine whether cognitive agility or the number of L2s the participants know affect the outcome in the first exposure task (LDT).

Study 2 more closely evaluates the effect of input (WR) by comparing two groups: the experimental group received the minimal exposure used in study 1, while the control group received no input. The comparison of two specific age groups allows for a more detailed investigation of the age effect investigated in Study 1. In addition, by using a quasi-longitudinal design in Study 2, we are able to examine whether the observed learning effects remain after a one-week consolidation period.

## **3 A lifespan perspective (Study 1)**

### **3.1 Introduction and Background**

Previous studies (Gullberg et al., 2012; Gullberg et al., 2010; Roberts et al., 2010; Veroude et al., 2010) have shown that adults can break into a new language system at first contact. Roberts et al. (2010) found increased lexical decision performance in an adult group that received input compared to an adult group without any exposure. They concluded that participants had implicitly learned some abstract phonotactic rules of the new language from this brief exposure.

However, a number of additional questions await response. In Study 1, we therefore asked the following three main questions to further investigate the matter: First, how quickly can adults and children learn to distinguish sound regularities in natural language input? Second, how does the ability to break into a new language at first contact develop across the lifespan? Do children learn these things more easily than adults (or vice versa)? Third, is this ability influenced by cognitive variables such as intelligence and working memory or the number of languages a person knows? Does knowing more foreign languages make the task of learning another new language easier or harder?

#### **3.1.1 From language input to learning**

In order to examine how learners tackle the three different sub-tasks (cf. e.g. Christiansen et al., 1998; Klein, 1986) of comprehending the utterance, encoding statistical regularities, and integrating these regularities, we first need to make sure that all learners share the same prior experience and knowledge. Artificial and statistical language learning studies have solved this problem by controlling the language input. However, they usually use very small samples of a language and often train learners prior to the task, for example through repetition, to guide the segmentation process. This is hardly comparable to naturalistic L2 acquisition at first contact. A few recent studies (Carroll & Widjaja, 2013; Hayes-Harb, 2007; Shoemaker & Rast, 2013) have used natural language and trained participants on these stimuli. The question still remains how well adults perform without any prior training on natural language stimuli.

In a different strand of research, classroom studies have used naturalistic settings to examine L2 learning of natural language at first contact (e.g. McLaughlin et al., 2004; Muñoz, 2006; Rast, 2008; Shoemaker & Rast, 2013). Although natural language was used, it was pedagogically prepared to help the learners break into the system. But what can learners do without any assistance, and how well is the input to learners really controlled? To probe these issues, we used natural language at first exposure, like Gullberg et al. (2010), Roberts et al. (2010), Veroude et al. (2010) and Gullberg et al. (2012), without any prior training or pedagogical help for the learners.

### **3.1.2 Age and Cognition**

Discussions about age effects often focus on ultimate attainment, ‘end states’, and native-likeness instead of on the process of development or the rate of attainment (see Birdsong, 2006 for overview). Supporters of the CPH have tried to find maturational ‘end points’ to explain an assumed ‘change in learning strategies’ in children’s relative to adults’ L2 acquisition (cf. e.g. DeKeyser, 2003), which is often used as an argument in favour of early foreign language learning (e.g. Abrahamsson & Hyltenstam, 2009). Adversaries of the CPH, on the other hand, question the claim that adults are not able to master L2 acquisition and support the notion that it is the amount of time spent learning a language that matters rather than the starting age (cf. also Carroll & Widjaja, 2013 mentioned above).

In a time of growing multilingualism, we probably need to reconsider the focus on nativelikeness and rather examine in more detail the importance of other skills, such as working memory and executive control processes required for language switching (e.g. Abutalebi, 2008; Abutalebi & Green, 2007; Adank & Janse, 2010; Bialystok et al., 2004; Hernandez, Martinez, & Kohnert, 2000). Multilingualism and globalisation also make the study of a broader age-spectrum increasingly relevant (Birdsong & Gertken, 2013; cf. the notions of multi-competence by Cook, 1992; Klein, 1998; the bilingual turn by Ortega, 2013). In much recent work, the influence of cognitive and social maturity is considered to exert rather positive than negative influence on language learning.

This study contributes new information in the following ways: Firstly, we examine participants across almost the whole lifespan. Secondly, in order to capture the very initial state of learning at first exposure, we test the ability to implicitly acquire (i.e. without instruction and directed attention) phonotactic information after only seven

minutes. Thirdly, we use continuous natural audiovisual speech to mimic the real life situation as closely as possible.

### **3.1.3 Research questions**

- 1) How quickly can adults and children learn to distinguish sound regularities in natural language input at first exposure without help? Can they generalize acquired knowledge from the input to new stimuli?
- 2) If learners can extract abstract phonotactic knowledge at all, does this ability change across the lifespan?
- 3) In what way is this ability influenced by cognitive variables such as intelligence and working memory? Does multilingualism help to grasp the phonotactic structure of an unknown foreign language?

## 3.2 Methods

### 3.2.1 Participants

For Study 1, we recruited and tested 168 participants (91 women, 77 men) between 10 to 90 years of age (out of 400 screened)<sup>2</sup>. Of these, 152 (84 women, 68 men) were retained for analysis. Since we treated age as a continuous variable, we ensured that participants were equally distributed across the following age bands: 10-12, 15-16, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79 and 80 plus year olds. For the interval from 10 to 20 we deliberately decided to introduce the bands 10-11 and 15-16<sup>3</sup>. No participant was a language expert (see below how we controlled this). All participants provided written consent, which, in case of children, was provided by their parents.

Participant selection proceeded in two steps. Participants first filled in a screening questionnaire. Participants who met the selection criteria then filled in a second language background questionnaire. Both questionnaires could be administered online<sup>4</sup>. The screening questionnaire only lasted five minutes and ensured the critical selection criteria for this study:

- 1) All selected participants spoke Swiss-German as their first language and Standard High German as their first second language;
- 2) All had at least a minimal understanding of English varying up to very good knowledge of English;
- 3) All had absolutely no knowledge of Chinese, Japanese, Thai or Swedish<sup>5</sup>;
- 4) All knew how to use a computer;
- 5) No one had any hearing or seeing loss and no corrective device; and
- 6) No one was a language expert. This last question was examined by an open ambiguous question where we asked participants to indicate with ‘yes’ or ‘no’ whether they ‘engage in the matter language in their daily lives’. If people said ‘yes’ they had to specify ‘how’. Through this specification we

---

<sup>2</sup> Participants were recruited as part of a bigger enterprise and tested on several tasks for other purposes (Project ‘A’ and Project ‘C’ of the Sinergia project; for more details cf. upcoming (projected for June 2014) *Bulletin Suisse de linguistique appliquée*, Vol.99, Issue1).

<sup>3</sup> To cover the whole continuum adequately, we would have had to recruit at least 10 people per numerical age – an objective for which we did not have enough resources.

<sup>4</sup> [www.soscisurvey.de](http://www.soscisurvey.de)

<sup>5</sup> Exclusion criteria for Project ‘A’

were able to judge whether the person really worked with language on a meta-cognitive plane in their daily lives (e.g. language teachers or linguists).

The screening questionnaire also provided information on gender and age as well as professional field. An initial attempt to control for socio-economic status was operationalised as having at least the Swiss Federal matriculation as an academic degree. In the end, we dropped this criterion for two reasons. It could not be applied to the youngest group and it was too selective for the elderly since only around 10% of the Swiss population acquired this degree before the 1980-ies (Bundesamt für Statistik (BFS), 2010).

Participants who met the criteria in the screening questionnaire were asked to fill out the second language background questionnaire prior to the actual experiment. The language background questionnaire lasted about 20 minutes and asked the following seven questions:

- 1) “How strongly are you interested in language(s) on a scale of 5 (not at all – strongly)?”
- 2) “Please rate your listening and reading comprehension of all the languages and dialects that you know according to the Common European Framework of Reference for Languages (CEFR), as outlined in the table below.<sup>6</sup>”
- 3) “Please indicate how you acquired the languages and dialects (except for Swiss-German) that you know: through school, a language course, an exchange abroad, in direct contact with natives or through media?”
- 4) “Does learning a new language come naturally to you, on a scale of 5 (not at all – very naturally)?”
- 5) “Which of these fields comes most natural to you (Grammar, Vocabulary, Pronunciation)? – Please put them in order.”
- 6) “How often do you listen, speak, write or read your languages in an average month on a scale of 5 (almost never – very frequently)?”
- 7) “How much do you like using foreign languages in general, on a scale of 5 (very reluctantly – very much)?”

Table 1 summarises participants’ knowledge of other ‘second’ languages (L2s). From here on, we will refer to any languages additional to the first language as L2s.

---

<sup>6</sup> Council of Europe (2011): CEFR codes: A1 (ab initio), A2, B1, B2, C1, or C2 (proficient)

Table 1: Participant characteristics in terms of the distribution of gender, age, number of foreign languages (Mean #L2s) and which foreign languages (L2s) were spoken by how many participants in each age group,  $n = 152$ .

Age	Total	Male	Female	Mean Age	Mean # L2s	L2s
10–12	21	12	9	10.6	2	English (20), French (13), Italian (2), Portuguese (1), Tamil (1)
14–16	19	8	11	15.4	3	English (19), French (19), Italian (2), Spanish (9), Latin (3), Portuguese (1), Greek (1)
20–29	16	6	10	25.7	4	English (16), French (16), Italian (10), Spanish (11), Portuguese (1), Serbian (1), Arabic (1), Telugu (1)
30–39	19	5	14	33.6	4	English (19), French (19), Italian (10), Spanish (8), Hungarian (2), Tagalog (2), Cebuano (2), Portuguese (1), Serbian (1), Czech (1), Swahili (1)
40–49	19	3	16	43.8	3	English (19), French (19), Italian (14), Spanish (2), Rhaeto-Romance (1), Hungarian (1), Sign-language (1)
50–59	17	8	9	54.9	3	English (17), French (17), Italian (10), Spanish (5), Russian (1), Portuguese (1), Tagalog (1)
60–69	17	7	10	64.6	3	English (17), French (16), Italian (13), Spanish (8), Latin (1), Portuguese (1), Greek (1), Hebrew (1), Rumanian (1)
70–79	20	16	4	72.5	3	English (20), French (18), Italian (14), Spanish (6), Portuguese (1), Rhaeto-Romance (1)
80+	4	3	1	83	3.25	English (4), French (4), Italian (3), Spanish (2)
Total	152	68	84	45	3	English (151), French (141), Italian (78), Spanish (51), Portuguese (7), Latin (4), Tagalog (3), Hungarian (3), Rhaeto-Romance (2), Russian (2), Greek (2), Cebuano (2), Serbian (2), Arabic (1), Czech (1), Rumanian (1), Tamil (1), Telugu (1), Swahili (1), sign language (1)

Most participants knew an additional foreign language (to Standard High German and English). Chinese was at least the fourth foreign language that participants

started to ‘learn’ (through our experiment) in terms of chronology. The information from the second questionnaire will not be evaluated further in the studies presented here.

Participants were recruited through e-mail and by word of mouth and were paid 150 CHF for their participation in the two questionnaires and the nine experimental tasks. Participants coming from cities other than Bern, Fribourg or Zurich were also reimbursed for travel.

### 3.2.2 Experimental Materials

#### *Exposure: Weather Report*

We used an carefully constructed audio-visual Weather Report (WR) to simulate minimal exposure to an unknown language (cf. Gullberg et al., 2010). The target language, Mandarin Chinese was unknown to the participants and typologically and genetically unrelated to their L1 Swiss-German. The Weather Report text consists of 120 coherent natural, but fully controlled clauses of Mandarin Chinese. Information structure principles of coherent discourse were respected as far as possible (see Appendix A for the complete text, written in pin yin with gloss and translation). The text consists of 292 different word types. On average, there were eight syllables per clause ( $M = 7.85$ , range 4-15). All words were controlled for frequency and tone. The Weather Report was spoken by a female native speaker who read the text in Chinese characters off a tele-prompter. The film was constructed to be as authentic as possible. Six weather charts were shown for different regions of an imaginary country. From time to time, the speaker highlighted what she was saying by pointing at the weather chart with a gesture. The occurrence of the gestures was scripted and controlled for to occur with certain words. An example of the Weather Report is shown in Figure 1. In total, the film lasted seven minutes.

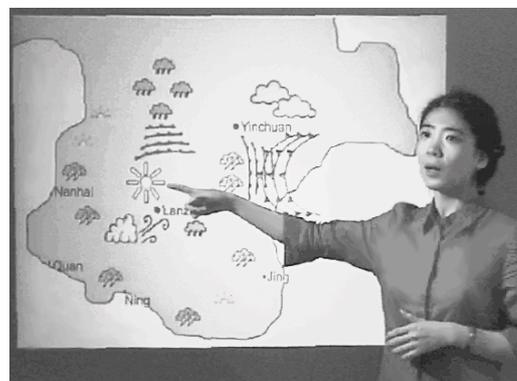


Figure 1: A video-frame from the Weather Report

### ***Experimental testing: Mandarin Chinese Lexical Decision Task***

We tested acquired phonotactic knowledge from the Weather Report using a Lexical Decision Task (LDT). In the LDT, participants have to decide whether a presented word is Chinese or not. The materials consisted of 256 monosyllables, half of which were real Chinese words and served as filler items, and half of which were experimental items (see Appendix B for the materials). Of the 128 real Chinese words, 64<sup>7</sup> had appeared in the Weather Report, while the other 64 as well as the 128 non-words were heard for the first time in the LDT. The 128 experimental items contained phonotactic violations of four different types (see Table 2).

Table 2: Four violation types of the 128 non-words used in the Lexical Decision Task

	# of items	Consonant cluster	Example	Consonant cluster	Example
Non-words	32	<u>CCCV</u>	<i>schra</i>	<u>VCCC</u>	<i>alst</i>
Non-words	32	<u>CCV</u>	<i>sna</i>	<u>VCC</u>	<i>ans</i>
Non-words	32	<u>CVC</u> (illegal nasal)	<i>gam</i>	<u>CVC</u> (illegal plosive)	<i>mat</i>
Pseudo-words	32	Phonotactically correct, but non-existent in Chinese			<i>chueng</i>

(Real) Mandarin Chinese contains about 400 syllabic types and a simple syllable structure: the onset can contain only one consonant (or none), and the coda can contain only one of two nasals (or none) (Bassetti, 2006). In other words, word-initial or word-final consonant clusters are not allowed and a simple CV-structure is preferred (Chen, 1991). On the basis of this, we constructed four types of phonotactic violations: three-, and two-consonant clusters, CVC syllables ending in an illegal consonant, and CVC syllables that are phonotactically correct but non-existent in Mandarin Chinese (pseudo-words). We assumed that it would be easy to classify illegal three-consonant clusters as non-words. Furthermore, we also assumed that it would be more difficult to classify illegal two-consonant clusters and even more difficult to classify CVC syllables. The responses to the pseudo-words were expected to mirror those of real Chinese words. Since we were not interested in the real Chinese words in our experiments (they served as filler items), we ignored the pseudo-words.

The CVC syllables are the experimental items we are most interested in. CVC syllables ending in an illegal plosive (CV\_plosive) were supposed to be easier to

---

<sup>7</sup> The 64 real words that appeared in the WR are controlled for frequency: 16 appeared in the WR each once, 16 twice, 16 three times and 16 four times (for details see Appendix B).

reject as non-words than CVC syllables ending in an illegal nasal (CV\_nasal), since a plosive in the syllable offset does not occur in Mandarin Chinese at all. In ancient Chinese, word-final plosives /p/, /t/ and /k/ were possible, and while Cantonese has preserved them, Mandarin Chinese has dropped them completely (cf. Lee, 1976, p. 4). This constitutes one of the major differences between Cantonese and Mandarin and has lead naïve listeners to suggest that the former sounds more like German while the latter sounds more like French (ibid.). The three possible nasals in Mandarin, however, contrast in two places of articulation: alveolar /n/ is possible in the syllable onset and coda position, while bilabial /m/ is only allowed in the syllable-initial position and velar /ŋ/ is only allowed in the syllable-final position (Lai, 2012). Therefore, the correct classification of our violated CVC syllables is only possible if the participants ‘learn’ from the input that these specific syllables are not possible in Mandarin Chinese (compared to rejecting the whole syllable-cluster in the case of CCCV, VCCC, CCV and VCC).

All stimuli of the LDT were re-recorded for this experiment by the same speaker that spoke the syllable material for the study by Gullberg et al. (2010). We introduced a silence-wave of 250 miliseconds before and after each stimulus. The CCCV- and CCV-syllables sounded bi-syllabic because an epenthetic vowel was accidentally provided<sup>8</sup>. Due to this, the two syllable conditions (CCCV and CCV) were removed from the analysis, because they could not be used as the three- and two-consonant-clusters. This left us with four non-word syllable conditions that all contain illegal consonants only in the syllable offset (VCCC, VCC, CV\_nasal, CV\_plosive).

A left-handed version of the experiment was computed to rule out a potential reaction time bias for handedness. The two randomized fixed orders from Gullberg et al. (2010) (List A versus List B) were kept to control for position effects. We decided to leave out the 32 bi-syllabic fillers items that were used in other experiments for additional purposes that were not of interest here. We also changed the inter-stimulus-interval (ISI; here: the time after the given response until the beginning of the next stimulus) from 1500 to 750ms<sup>9</sup> to shorten the experiment. At

---

<sup>8</sup> It is possible that our speaker accidentally provided a vowel in the syllable-initial consonant clusters because Mandarin does not allow consonant clusters and pronouncing such clusters with Chinese intonation poses a difficult task per se (cf. Dupoux et al., 2001).

<sup>9</sup> The 750ms are divided into 200ms before the appearance of the fixation star (which was done after reports in pilots where participants interpreted the appearance of the fixation star as the confirmation of their answer instead of the announcement of the next stimulus), 300ms

the same time, the maximum response time was increased from 1500 to 2050ms<sup>10</sup> to allow for the possibility that elderly participants' responses may be considerably longer.

The two measured outcome variables of the LDT were accuracy and reaction time. Accuracy was coded as '1' for both correct hits and correct rejections and '0' for both misses and false alarms. We added five practice trials<sup>11</sup> before the beginning of the experiment to unify the task with the other group-projects<sup>12</sup>.

### 3.2.3 Cognitive Tests: additional Materials

#### *Crystallized and Fluid Intelligence (Gc & Gf)*

As a measure of crystallized intelligence, we used a lexical test in the participants' first L2, Standard High German, the so-called 'Wortschatztest' (WST, Schmidt & Metzler, 1992). In this test, 42 lines with increasingly more difficult words are presented. The participants' task is to find the correct word in each line, between five non-words.

We used the Raven matrices as a measure of fluid intelligence (Raven, 1962). This test contains five sets (A, B, C, D, E) of 12 items/pictures with a missing piece each. We used sets C, D, and E with eight possible pieces each to complete a total of 36 pictures, arranged in order of difficulty. We used the advanced matrices (designed for adults) for all participants (including the children) to ensure a comparable measure for the whole group. Scores are given for each correct answer and are converted into standardized ranks through tables based on the participants' age.

For the development of crystallized intelligence with age, we expected a (simplified)<sup>13</sup> total linear increase across the age span (cf. Cattell, 1987).

---

duration for the appearance of the fixation star, and a 250ms silence built into the beginning of each stimulus.

<sup>10</sup> 1800ms instead of 1500ms, plus the silent wave after the stimulus, 250ms = 2050ms.

<sup>11</sup> For the practice trials, we used five of the removed bisyllabic filler items.

<sup>12</sup> Project 'A' and Project 'C' both used practice trials before their experimental task; for detailed description of these projects compare upcoming Vol.99, Issue 1 of Bulletin Suisse de linguistique appliquée (projected for June 2014).

<sup>13</sup> We are aware that the correlation of these two variables (Gc and Gf) along the lifespan would be more of a curvilinear correlation (cf. e.g. Baltes, 1987) and the current study is not meant to question this relationship. However, we are more interested in the general trend, which is why the linear relationship should suffice to illustrate this point.

Correlations between accuracy in the experimental task and crystallized intelligence were expected to be weakly positive to possibly negative (cf. e.g. Robinson, 2010), since the experimental task was an implicit learning tasks. For the development of fluid intelligence with age, we expected a (simplified) general linear decline across the age span (cf. Cattell, 1987). A tendency for better performance on our experimental task with better fluid intelligence scores was expected (cf. e.g. Robinson, 2010).

### ***Working memory (WM)***

We chose our WM task such that the influence of the experimenter is minimal. In reading span tasks, for example, participants are asked to repeat words that have to be written down by the experimenter (Daneman & Carpenter, 1980; cf. also Mackey et al., 2002). In our experiment, the problem of inter-rater variability would have arisen using this task, because four different experimenters conducted the experiments. Hence we used the backward digit span task from the Wechsler Intelligence Test for Adults (HAWIE, Tewes, 1991). Since our implicit learning task is audio-visual and does not involve reading, the digit span task suits our purpose well. Therefore, we recorded the number-strings that have to be repeated by the participants, in order to ensure inter-rater reliability. Number-strings start from two-digit numbers and increase up to maximally eight-digit numbers (=7 'levels'). There are two number-strings per 'level'. The task lasts as long as the participant repeats at least one number-string per 'level' correctly (in reverse order). There are two results: the digit span that measures the 'level' the participant has reached, and the total of correct responses.

We expected positive correlation between WM and age for children and a negative one for adults (cf. also Borella et al., 2008; Gathercole et al., 2004; Jenkins et al., 1999; Park et al., 2002; Park & Payer, 2006; Salthouse, 1994, 1996), as well as positive correlations of WM with accuracy in our experimental implicit learning task (cf. e.g. Gathercole et al., 2004; Salthouse, 1994).

### ***Executive Control***

We used the Simon task as a measure for executive control processes. The comparison of children's and adults' executive control processes requires tasks with different degrees of difficulty given the different developmental stages of the prefrontal cortex (cf. e.g. Huttenlocher & Dabholkar, 1997), which the Simon task fulfills (see review in Lu & Proctor, 1995). The task was based on the version used by Bialystok, Craik, Klein and Viswanathan (2004), but translated into Standard

High German<sup>14</sup>. It consisted of 28 trials, during which a red or a blue square appeared either on the right or the left side of the computer screen. There were two types of trials, ‘congruent’ and ‘incongruent’. A trial was ‘congruent’ if the red square appeared on the right side or if the blue square appeared on the left side of the screen. In ‘incongruent’ trials, the squares appeared in reverse positions: red on the left side and blue on the right side. There was an equal amount of congruent and incongruent trials (14 each). Congruent trials are supposed to yield faster reaction times; reaction times are measured from the onset of the stimulus till the given response. The positive difference between congruent and incongruent trials is called the Simon effect (congruent trials are usually faster and more accurate).

Ellis and Sagarra (2010) have suggested that executive control processes may play a role in what gets processed and thereby contribute to the final level of L2 attainment. We therefore expected a positive correlation with executive control (measured by the Simon Effect) and increasing age, as well as a positive correlation of more executive control and accuracy in our experimental paradigm.

#### ***Number of L2s (# L2s)***

Increasing evidence indicates a bilingual or multilingual advantage, assuming that bilinguals (or multilinguals) frequently switch between languages, which enhances executive flexibility (e.g. Bialystok, 2011; Bialystok, Craik, et al., 2005; Bialystok et al., 2004; Bialystok et al., 2008; Bialystok, Martin, & Viswanathan, 2005; Craik et al., 2010) and might even delay the onset of Alzheimer disease (Craik et al., 2010). Kavé et al. (2008) have shown that the number of languages known significantly interacts with outcomes on cognitive measures, discussing their results in the context of ‘cognitive reserve’ theories. As a result of these findings and with the notions of multi-competence in mind, we expected a positive correlation between a wider foreign language repertoire and the acquisition of an unknown language.

#### ***German Lexical Decision Task (LDT baseline)***

A German Lexical Decision Task was administered as a baseline for reaction times. The task was mainly used for the other sub-projects<sup>15</sup> in the bigger enterprise, and will not be further analyzed as part of this study. It consisted of 20 words and 20 non-words. Non-words were similar but not identical to words with regard to

---

<sup>14</sup> The translation was done by Jan Vanhove (2014).

<sup>15</sup> For detailed description of these projects compare upcoming Vol.99, Issue 1 of Bulletin Suisse de linguistique appliquée (projected for June 2014).

number of syllables. They were created by changing one or two phones in existing German words. All non-words ('pseudo-words') were rated by three German speakers as phonotactically possible but non-existent in both Standard High German and Swiss German. A male speaker who works as a professional radio announcer with the German-language national radio spoke the stimuli in the Swiss variant of Standard High German ('Schweizer Hochdeutsch'; cf. Vanhove, 2014 for more details of the task).

### ***English Proficiency Test***

A 20-item multiple choice grammar test (Allen, 1992) and a 25-item English C-test<sup>16</sup> were administered as a comparison to the self-evaluation scores of the English proficiency levels that participants indicated in the second background questionnaire. This measure mainly served sub-project 'C' and will not be further investigated within the scope of this study.

### **3.2.4 Procedure**

The data collection for this study was interlaced with data collections from two other studies (Project 'A' and Project 'C' of the Sinergia project). This not only made the test sessions long, but also affected test situations in various ways.

### ***Data acquisition***

Participants were mainly tested at three different locations: the Universities of Bern, Fribourg and Zurich. At these locations, a quiet room was provided. Participants were always asked to switch off their mobile phones. Sometimes, more than one participant was tested at the same time. Up to four participants could be comfortably seated in the same room, back to back. In cases of multiple testing, participants were instructed together at the beginning and asked to whisper if they had additional questions during the experiment as well as to quietly leave the room during the break. During multiple testing, there were at least two examiners present to ensure prompt supervision and to avoid eye contact and mutual distraction between participants. In some cases, we had to travel to people's homes in order for them to participate in our experiment. The reasons for this were limited possibilities of locomotion either due to age or childcare. In one case, we went to test at a local school where 12 children took part in three slots on the same day. Two classrooms were offered to us so we could carry out the experiments without disturbances.

---

<sup>16</sup> <http://www.sprachenzentrum.uni-rostock.de/einstufungstests/c-test/c-test-englisch/>

Despite these variations in data collection circumstances, we made all efforts to maintain similar test conditions.

### ***Experimental tasks***

Table 3 shows the sequence of the experimental tasks. The entire session consisted of ten tasks with two breaks in-between. The order of the tasks in the three sub-projects of the Sinergia project ('A', 'C', 'D') was randomized, while the rest of the tasks were always presented in the same order to control for fatigue effects. Our study is referred to as sub-project 'D' in the table and always implies the presentation of the Weather Report (WR) followed by the execution of the Lexical Decision Task (LDT).

The WR was presented on a laptop screen and participants were asked to listen to the sound over headphones with no other instructions ("please, just watch this film") in order to promote implicit learning without focusing attention to the phonotactic structure of the unknown language. The LDT was presented right after the exposure to the WR.

The experiment was programmed in E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). We used a Cedrus button-box RB-834 as input device for the classification of words into 'Chinese' or 'not Chinese'. The right button was marked 'O', and the left one 'X'. Participants were wearing headphones MBK C 800 throughout our experiment. Instructions were presented visually, in Standard High German, while stimuli appeared only auditorily, separated by a fixation star that appeared after maximally 2050ms of non-response or 200ms after the given response. Participants were told to answer as accurately and quickly as possible and to answer before the appearance of the fixation star. There was a 750 milliseconds break after the participant gave the answer and the beginning of the next trial. Five practice trials preceded the actual LDT. Participants had the chance to repeat the practice trials if desired and/or could ask questions in case of any uncertainties. No more questions were allowed once participants started the experiment. This applied to all experimental and cognitive tasks.

### ***Gc-test***

The Wortschatz-Test was carried out as a paper and pencil task. The instruction was written on the front page in Standard High German. There was no time restriction for this task.

### ***German LDT***

The German Lexical Decision Task was programmed in E-Prime 2.0 (Schneider et al., 2002). Participants were told that they were about to be presented with a series of stimuli, some of which were existing German words whereas others were made up. They were instructed to press the right button on the response pad (marked 'O') if they thought the stimulus in question was an existing word and to press the left button (marked 'X') if they thought it was not, or vice versa for left-handed participants (using Cedrus button-box RB-834 and headphones MBK C 800). We asked them to make their decisions as accurately and as quickly as possible. The order of stimuli was kept constant for all participants. The answering time for each stimulus was unlimited and the time after the given answer to the beginning of the next trial was 1500 milliseconds. Each trial started with a 1000-millisecond fixation period during which a focus point ('o') was displayed in the centre of the screen. The focus point remained on-screen during stimulus playback. Five practice trials preceded the actual LDT that had to be answered correctly in order to proceed to the actual experiment.

### ***English test***

The English-Test was carried out as a paper and pencil task with a simple instruction ("please fill in the gaps"). There was no time restriction for this task.

### ***Simon task***

The Simon Task was programmed in E-Prime 2.0 (Schneider et al., 2002). Participants were told that they should press the button on the right side in case the red square appeared on the screen, and the left button for the blue square, using Cedrus button-box RB-834. Responses should be made as quickly and as accurately as possible. The task began with eight practice trials that had to be successfully (= without any mistakes) completed before proceeding to the experimental trials. Almost all our participants repeated the practice trials at least once. The 28 experimental trials, half of which were congruent trials and half incongruent ones, were presented in randomized order.

### ***WM-test***

In the Backward Digit Span Task, participants were told that they would hear number-strings of varying lengths only once, and were instructed to list them in reverse order to the experimenter. The experimenter then read an example number-string of three digits off a standardized instruction sheet. If participants reversed the sequence correctly, the experimenter proceeded directly to the actual experiment. If

not, participants were corrected and presented with a second example until the task was fully understood. Participants were informed that they would not receive any feedback until the end of the task. The number-strings were played from a recording, using Praat software (instead of spoken by the experimenter as foreseen in the original task). The experimenter played number-strings and directly rated the answers onto the instruction sheet in a way that participants could not see what the experimenter was writing. The task was finished as soon as participants made two mistakes on the same level.

### ***Gf-test***

In the Raven Task, participants were asked to solve as many pictures as possible without any time pressure (they could hand in their solutions without completing all pictures). Time allowance was set to a maximum of 60 minutes. The task was administered as a paper and pencil task.

### ***Overview***

Table 3: Experimental design of ten tasks

<b>Experimental task</b>	<b>Specification</b>	<b>Time</b>
<i>Introduction</i>	<i>orally</i>	3-5'
WST ('Wortschatztest') (Gc)	Paper and pencil	5-20'
German Lexical Decision Task	PC, headphones, button-box	5-10'
Sub-project A, C or D	PC, headphones and either mouse, keyboard, or button-box	10-20'
Sub-project A, C or D	PC, headphones and either mouse, keyboard, or button-box	10-20'
<i>Break</i>	<i>- drinks and refreshments -</i>	<i>10'</i>
Sub-project A, C or D	PC, headphones and either mouse, keyboard, or button-box	10-20'
Sub-project A, C or D	PC, headphones and either mouse, keyboard, or button-box	10-20'
<i>Break</i>	<i>- drinks and refreshments -</i>	<i>20'</i>
English C Test	Paper and pencil	10-20'
Simon Task	PC, button-box	3-5'
Backward digit span task (WM)	Paper and pencil, headphones, audio-files played via Praat	3-5'
Raven Matrices (Gf)	Paper and pencil	10-60'
<i>Debriefing</i>	<i>orally</i>	<i>3-5'</i>
<b>10 tasks</b>		<b>122'-240'</b>

In sum, our experimental design consisted of ten tasks<sup>17</sup> (cf. Table 3). A first break of about ten minutes was allowed after four experimental tasks and the second break of 20 minutes after another two tasks. During the breaks, drinks and snacks were offered to the participants and they were recommended to stand up and walk around

<sup>17</sup> Sub-project ,C' consisted of two sub-tasks.

the room (or leave the room in case of multiple testing). After the experiments, participants were informally asked if they made use of any meta-linguistic strategies (i.e. “What did you focus on when you listened to the Weather Report? Did you use a strategy to decide whether a word was Chinese or not?”).

### 3.3 Data Analysis

#### 3.3.1 Data cleaning

Initial data inspection involved the identification of participant outliers. We excluded eight participants because of zero correct responses in any syllable condition of our experimental task and eight additional participants because of missing data in the cognitive variables. This left us with 152 out of 168 tested participants (91%).

Four more steps of data cleaning and transformation were done:

- Reaction times smaller than 100ms and bigger than 2000ms as well as trials with an onset delay bigger than 250ms were removed since these answers constituted technical errors. This step affected about 2% of all trials.
- Log-transformation of reaction times. Our methods for data analysis assume normality of the data, but the empirical distribution was slightly left-skewed (see left panel in Figure 2). The log-transformation stretches out small values and squeezes the bigger ones thus restoring the assumption of normality on the transformed data (see right panel in Figure 2).
- Closer analysis of reaction time outliers was done according to Osborne (2013), Baayen and Milin (2010), and Ratcliff (1993), using winzorizing for reaction times above or below two standard deviations of the group mean of the specific syllable conditions. This procedure affected on average 2.3% of the data per syllable condition.
- Accuracy was initially coded as ‘1’ for both correct hits and correct rejections and ‘0’ for both false alarms and misses. The proportions of correct answers (= correct hits plus correct rejections) were then transformed by arcsine-square-root. As with reaction times, original values violate the assumption of normality, while the transformed values satisfy it. This procedure is widely used when dealing with proportions (cf. e.g. Howell, 2002; Osborne, 2013).

No further data cleaning was undertaken (and importantly, no winzorizing or trimming of the accuracy data).

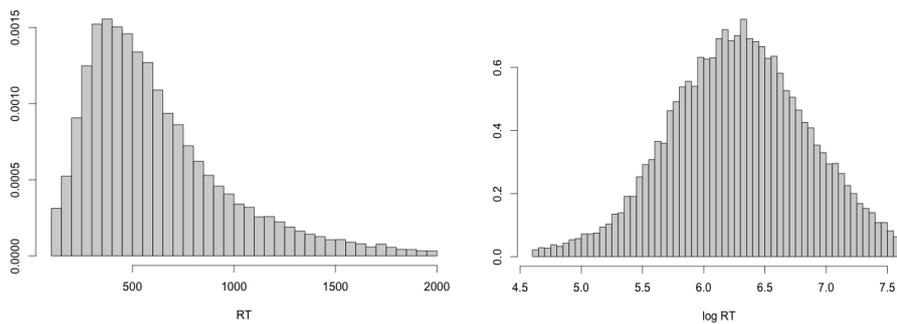


Figure 2: left panel: raw reaction times across all subjects, across all conditions after removal of responses <100ms and >2000ms; right panel: log-transformed reaction times.

### 3.3.2 Statistical analyses

The results from the Lexical Decision Task (LDT) were analyzed in terms of accuracy and reaction times. We started by comparing original (untransformed) mean proportions of hit rate and reaction times. Next, hit rate was arcsine-square-root-transformed, reaction times were log-transformed and their mean values across all ages of the four non-word conditions (VCCC, VCC, CV\_plosive and CV\_nasal) were compared. Only the reaction times of the correct answers were considered. The real words served as filler items. Of the 128 filler items, 64 have appeared in the Weather Report and were not investigated further. The 64 filler items that were heard for the first time in the LDT served as a reference category to compare the overall accuracy on words and non-words<sup>18</sup> (non-words were all heard for the first time in the LDT). The only analysis carried out with the filler items was, first, a t-test to control that fillers were responded to equally correctly as the experimental items. A second t-test examined whether the mean response accuracies of filler items differed significantly from chance. Individual t-tests were carried out for the filler items as well as for each of the four non-word conditions.

Next, an ANOVA examined whether the means of the four non-word conditions differed significantly, both in terms of mean accuracy (arcsine-square-root-transformed) and in terms of mean (log-transformed) reaction times of correct trials. Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007). Tukey's test corrects for experiment-wise error rate to control for the probability of

---

<sup>18</sup> Since we only analyzed non-word conditions VCCC, VCC, CV\_nasal and CV\_plosive, the number of total non-words also added up to 64 non-words.

an increased Type-I error rate and is therefore suitable for multiple comparisons. Tukey's test makes the following two assumptions: 1) the observations being tested are independent; 2) there is equal within-group variance across the conditions associated with each mean in the test (homogeneity of variance). If 0 is within the confidence interval [conf\_left, conf\_right], one cannot reject the hypothesis H0 that the true difference is zero. Our observations are independent insofar as 152 different participants were tested and each solved the LDT only once. However, strictly speaking, since we treated age as a continuum and did not compare different age groups, the four syllable conditions served as groups, which are not completely independent as every participant solved all items in each of the four syllable conditions. Nevertheless, we decided to use Tukey's test since the number of 152 participants is sufficiently large to mimic independence. The second assumption is met by the confirmed Levene's test. The effect size of the post-hoc multiple comparisons was reported with Cohen's d (Cohen, 1988).

In a next step, we analyzed how age of a participant influenced the hit rate and the mean reaction time given a syllable condition. K-fold cross-validation was used for assessing which model – linear or quadratic – would best predict the data. At each round, data was uniformly at random split in two sets, a 'training' and a 'testing' set. Model parameters were estimated on the 'training' set first, followed by an evaluation on the remaining 'testing' set. Multiple rounds of this cross-validation process were run in order to reduce variability, and the resulting validations were averaged across the rounds. We used 10 rounds with 90% of the data for training and 10% for testing. The sum of the squared residuals was used to measure the error. We compared mean cross-validation errors and corresponding standard deviations across the models and visually inspected how they fitted the data (see Appendix C). Relatively slight differences in mean cross-validation errors given the high standard deviations did, in our opinion, not justify a more complex, quadratic model. Therefore, we examined possible correlations between the mean accuracy and reaction time scores per syllable condition and age using a Pearson Correlation.

In a last step, we investigated the correlations of the cognitive variables with age and with accuracy and reaction time in each syllable condition using Pearson Correlations. To simplify the measure of 'multilingualism', we only tested whether the number of foreign languages known ('Number of L2s') contributed to the prediction of our dependent variables. Crystallized and fluid intelligence test scores were also treated as numbers, as well as working memory scores from the digit span

task, where we only considered the total number of correct responses and did not analyze the digit span. For the Simon task, we analyzed the Simon effect only.

### 3.3.3 Hypotheses

- 1) Adults and children learn to distinguish sound regularities after 7 minutes.
  - a. VCCC syllables are easier to reject as non-words than VCC syllables
  - b. VCC syllables are easier to reject as non-words than CVC syllables
  - c. CV\_plosive syllables are easier to reject as non-words than CV\_nasal syllables
  - d. a., b., and c. can be seen both in terms of accuracy and in terms of reaction time.
- 2) Adults perform better than children.
  - a. In terms of accuracy;
  - b. But NOT in terms of reaction time.
- 3) Each cognitive variable (# L2s, Gc, Gf & WM) correlates with different strengths with the four syllable conditions: (strongest) VCCC > VCC > CV\_plosive > CV\_nasal (weakest correlation)

### 3.4 Results

For reasons discussed in the Method Section 3.2.2, we will report the results on four non-word syllable conditions only, namely the two ‘control’ conditions VCCC and VCC, and the two ‘critical’ conditions CV\_nasal and CV\_plosive (= 64 items; cf. Table 4 as a reminder of the four conditions). The results on the real words serve only as reference values for the results of the non-words. We therefore only refer to the real words that appeared for the first time in the LDT (= 64 items). If we refer to the non-words as compared to the real words, we always mean the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) taken together.

Table 4: The four non-word conditions in the LDT. The illegal consonants are underlined.

	Items	Consonant cluster	Example	Items	Consonant cluster	Example
‘Control’	16	<u>VCCC</u>	<i>alst</i>	16	<u>VCC</u>	<i>ans</i>
‘Critical’	16	<u>CVC</u> (illegal nasal)	<i>gam</i>	16	<u>CVC</u> (illegal plosive)	<i>mat</i>

#### 3.4.1 Mean accuracy and reaction times per syllable condition

We started by computing the mean hit rate (i.e. accuracy; correct rejections for non-words and correct hits for fillers) and mean reaction times in the LDT across all participants and ages. Table 5 summarises the untransformed mean accuracies and untransformed mean reaction times for non-words and real words (= fillers that were heard for the first time in the LDT).

Table 5: Mean and standard deviation (in brackets) of untransformed hit rate and reaction times (in ms) on the Lexical Decision Task per non-word syllable condition and real words.

	Mean hit rate (SD)	Mean reaction time (SD)
VCCC	.895 (.171)	470 (221)
VCC	.777 (.205)	540 (184)
CV_nasal	.496 (.227)	735 (234)
CV_plosive	.750 (.195)	701 (199)
Fillers	.707 (.124)	606 (155)

The resulting proportions were then transformed to arcsine-square-root values and the reaction times were log-transformed. Note that chance level would be equal to .50 in proportions, but equals .79 in arcsine-square-root. Figure 3 summarises the arcsine-square-root-transformed mean hit rates and the transformed mean reaction times by condition collapsed across all participants.

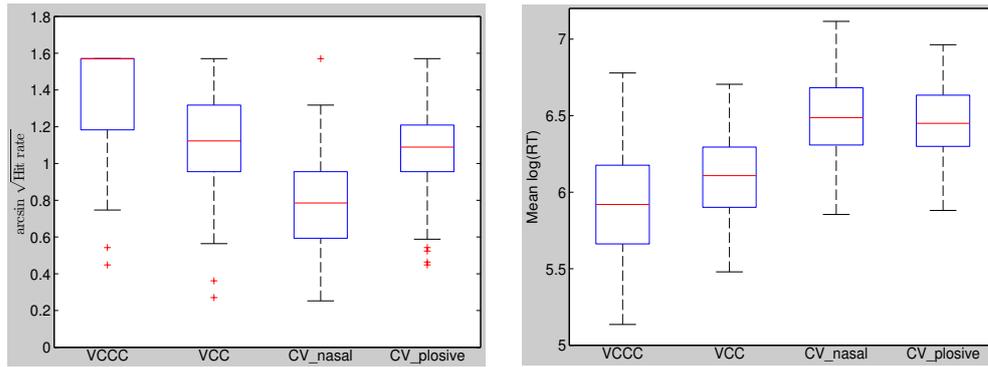


Figure 3: Boxplots comparing (left) the mean transformed hit rate and (right) the mean transformed reaction times, averaged across all participants in each syllable condition. Arcsine-square-root-transformed hit rate at the chance level equals 0.79,  $n = 152$ .

Figure 3 shows that, in terms of hit rate, phonotactically illegal syllables with three-consonant-clusters (VCCC) were overall easier to identify than illegal two-consonant-clusters (VCC). Two-consonant-clusters (VCC) were easier to identify than CV\_nasal syllables, and CV\_plosive syllables were easier to identify than CV\_nasal syllables. In terms of reaction time, phonotactically illegal syllables with the three-consonant-cluster were responded to more quickly than illegal two-consonant-cluster-, CV\_nasal and CV\_plosive syllables.

### 3.4.2 Comparison of syllable conditions

#### *Accuracy*

After Levene's test for equal variances was confirmed ( $F = 2.06$ ,  $p = .104$ ), a one-way Analysis of Variance (ANOVA)<sup>19</sup> was carried out with Syllable Condition (VCCC, VCC, CV\_nasal and CV\_plosive) as the independent, and Accuracy as the dependent variable on the transformed values (cf. Figure 3). There was a significant main effect for Syllable Condition ( $F(3, 604) = 110$ ,  $p = .00001$ ). In other words, the response accuracy varied for the different syllable conditions.

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007)<sup>20</sup>.

Table 6 summarises the results of the post-hoc multiple comparison. The analyses showed that each syllable condition was significantly different from every other syllable condition in terms of mean hit rate, except for CV\_plosive from VCC. The results thus confirm hypothesis 1a) that VCCC is easier to reject than VCC, and 1c)

<sup>19</sup> For all reported ANOVAs, alpha was set at 0.05. For ANOVA-tables cf. Appendix C.

<sup>20</sup> Cf. Method Summary section of Study 1 for description of Tukey's test.

that CV\_plosive is easier to reject than CV\_nasal. Hypothesis 1b) is only partly confirmed, because VCC is not easier to reject than CV\_plosive, but is easier to reject than CV\_nasal.

Table 6: Tukey's honestly significant difference criterion for mean transformed hit rate in the comparisons of the four conditions VCCC, VCC, CV\_plosive and CV\_nasal.

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
VCC - VCCC	-0.208	-0.289	-0.127	-0.723	1
VCC - CV_nasal	0.359	0.278	0.440	1.290	1
VCC - CV_plosive	0.053	-0.028	0.134	0.190	0
CV_plosive - CV_nasal	0.306	0.225	0.387	1.174	1
VCCC - CV_nasal	0.567	0.486	0.648	2.105	1
VCCC - CV_plosive	0.261	0.180	0.342	0.964	1

### **Reaction times**

A Levene's test for equal variances was not confirmed for the comparison of the four conditions ( $F = 6.17, p = .0004$ ). An analysis of variance (ANOVA) with Syllable Condition (VCCC, VCC, CV\_nasal and CV\_plosive) as the independent and Reaction Time as the dependent variable was conducted and revealed a main effect of Syllable Condition ( $F(3, 604) = 103, p = .00001$ ). In other words, the mean reaction times varied for the different syllable conditions.

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007).

Table 7: Tukey's honestly significant difference criterion for mean transformed reaction times in the comparisons of the four conditions VCCC, VCC, CV\_plosive and CV\_nasal.

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
VCC - VCCC	0.181	0.093	0.269	0.555	1
VCC - CV_nasal	-0.386	-0.474	-0.298	-1.344	1
VCC - CV_plosive	-0.357	-0.445	-0.269	-1.298	1
CV_plosive - CV_nasal	-0.029	-0.117	0.059	-0.107	0
VCCC - CV_nasal	-0.566	-0.654	-0.478	-1.768	1
VCCC - CV_plosive	-0.538	-0.626	-0.450	-1.736	1

Table 7 summarises the results of the post-hoc multiple comparisons. Again, the analyses showed that each syllable condition was significantly different from every other syllable condition in terms of transformed mean reaction times, except for CV\_plosive from CV\_nasal. The results thus confirm hypothesis 1a) that VCCC syllables are easier to reject as non-words than VCC syllables and 1b) that VCC syllables are easier to reject than CVC syllables. They do not support hypothesis 1c) since there was no difference between CV\_plosive and CV\_nasal.

### 3.4.3 Comparison of performance to chance level

In a third instance, we examined whether the mean response accuracies differed significantly from chance. Bonferroni corrected right-tailed Student's *t*-tests revealed that the accuracy scores for three of the syllable conditions were significantly different from chance, namely VCCC ( $t(151) = 25.0, p < .001$ ); VCC ( $t(151) = 14.9, p < .001$ ); and CV\_plosive ( $t(151) = 14.4, p < .001$ ). In contrast, responses to CV\_nasal syllables were at chance ( $t(151) = -.018, p > .05$ ). The results suggest that participants overall were able to correctly reject the consonant-cluster syllables as not being Chinese, and that they were also able to identify the CV\_plosives as not being Chinese. It is particularly noteworthy that illegal CV\_plosives were classified with accuracy significantly above chance (cf. Figure 3). This suggests that participants derived phontactic knowledge from the input, since the structure per se is possible in the participants' native language and in the target language, but the specific instantiation of CV\_plosive is only illegal in the target language. Moreover, as a group, they were guessing at the CV\_nasal syllables. The hit rate for non-words overall was significantly above chance ( $t(151) = 16.20, p < .001$ ), as well as the hit rate for real words ( $t(151) = 19.10, p < .001$ ), using arcsine-square-root values (cf. Figure 4).

### 3.4.4 Words compared to non-words

Next, we examined whether the mean transformed hit rate and the mean transformed reaction times of the real words differed significantly from the non-words. Figure 4 summarises the findings.

Bonferroni corrected right-tailed Student's *t*-tests revealed that the accuracy scores for the two word types did not differ significantly ( $t(151) = -1.57, p > .05$ ) and neither did the reaction times ( $t(151) = 2.08, p > .05$ ).

Figure 4 shows that neither the mean transformed hit rate nor the transformed mean reaction times for words and non-words differed significantly and that both conditions were answered significantly above chance level, overall (see above; arcsine-square-root-transformed chance level for hit rate equals 0.79). This suggests that participants successfully distinguished words from non-words.

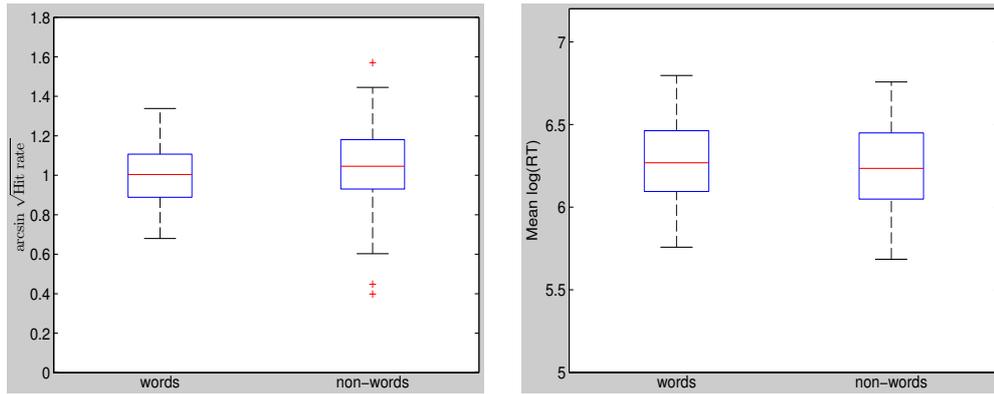


Figure 4: Boxplots comparing (left) the mean transformed hit rate and (right) the transformed mean reaction times, averaged across all participants in words and non-words. Arcsine-square-root-transformed hit rate at the chance level equals 0.79,  $n = 152$ .

### 3.4.5 Development across the lifespan

#### Accuracy

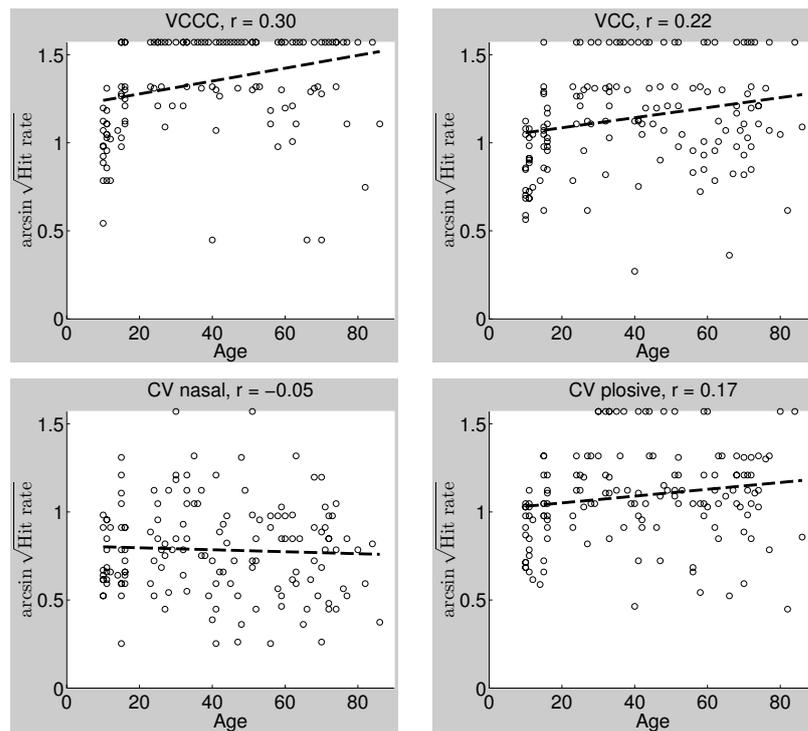


Figure 5: Linear models for transformed hit rate in the four syllable conditions and age. Transformed hit rate at the chance level equals 0.79,  $n = 152$ .

Next, we computed linear regression models for each syllable condition in terms of mean arcsine-square-root-transformed hit rate and age in order to examine the possible changes to responses across the lifespan. Figure 5 summarises the findings. The Pearson's correlations showed that age correlated significantly with responses to

VCCC and VCC ( $p < .001$ )<sup>21</sup>. Age also correlated significantly with CV\_plosive ( $p < .05$ ), but there was no correlation between the hit rate in CV\_nasal and age. According to Cohen (1988), the correlation coefficient was medium for condition VCCC ( $r(152) = .30$ ), for VCC ( $r(152) = .22$ ) and CV\_plosive ( $r(152) = .17$ ), and very small for CV\_nasal ( $r(152) = -.05$ ). Overall, in cases of trends of age, they went in the opposite direction from the Critical Period Hypothesis (CPH). That is, performance improved with increasing age or at least remained stable.

### Reaction times

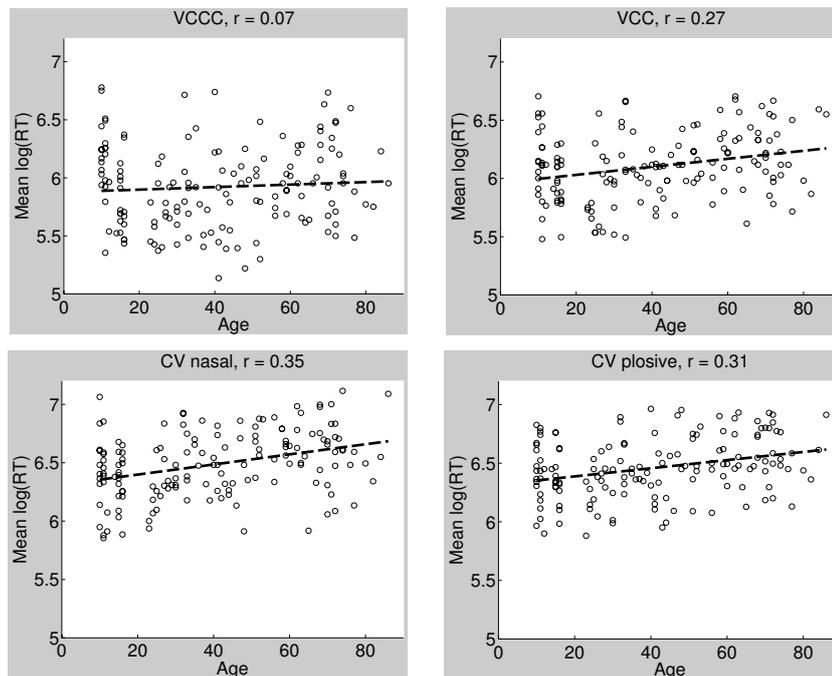


Figure 6: Linear models for mean transformed reaction times in the four syllable conditions and age,  $n = 152$ .

Figure 6 shows the linear model for each syllable condition in terms of mean transformed reaction times and age. The Pearson's correlations showed significant correlations between age and responses to CV\_nasal, CV\_plosive, and VCC ( $p < .001$ ), but there was no correlation for mean reaction times to VCCC and age. According to Cohen (1988), the correlation coefficient was medium for conditions CV\_nasal ( $r(152) = .35$ ) and CV\_plosive ( $r(152) = .31$ ), small for VCC ( $r(152) = .27$ ), and very small for VCCC ( $r(152) = .07$ ). There was a trend for higher mean reaction times with increasing age, which was expected from the literature (cf. e.g.

<sup>21</sup> The p-value calculator for correlation coefficients was used (Soper, 2013): <http://www.danielsoper.com/statcalc>; cf. also Cohen, Cohen, West and Aiken (2003).

Baltes, 1987). Furthermore, the pattern matched the accuracy data insofar as the CV\_nasal syllables were the hardest to reject and also took the longest to respond to.

### 3.4.6 Correlations with cognitive variables

In a next step, we performed analyses to investigate whether performance on the LDT correlated with measures of the cognitive variables. For the correlation of crystallized and fluid intelligence and working memory across the lifespan<sup>22</sup>, we expected developments according to the literature (cf. Baltes, 1987). For the correlation of each cognitive variable with the four syllable conditions, we expected the strongest correlations with syllable-condition VCCC, and VCC, and less strong correlations with CV\_plosive, and the weakest correlations with CV\_nasal.

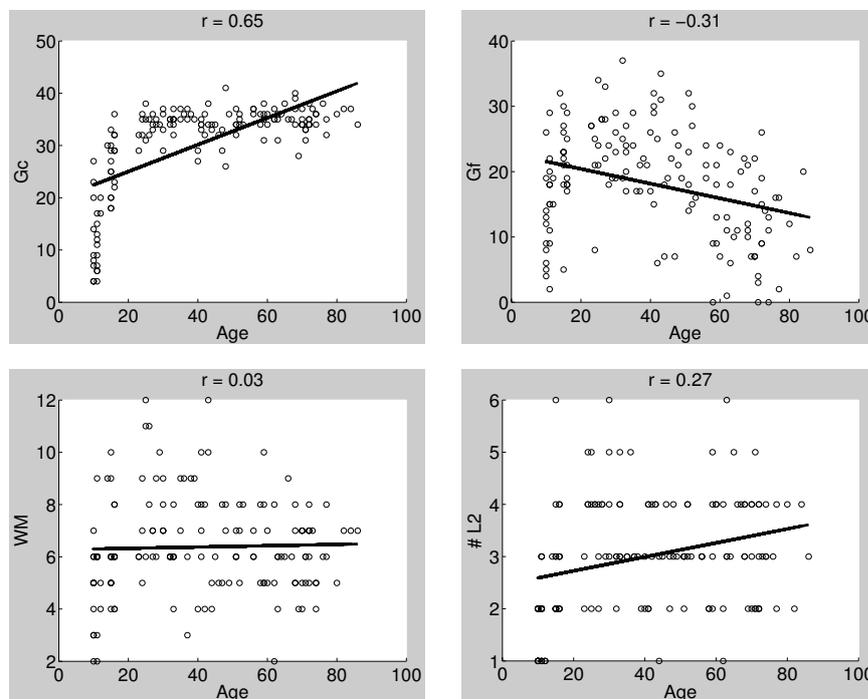


Figure 7: Linear correlation between age and cognitive variables. Top panel from left to right: crystallized intelligence (Gc), fluid intelligence (Gf); bottom panel from left to right: working memory (WM), number of foreign languages without German (#L2),  $n = 152$ .

<sup>22</sup> The results of the Simon task are not reported here, because we did not find a Simon effect (i.e. an increased time needed to respond to incongruent items) similar to that reported by Bialystok and colleagues (2004). One possible interpretation is that participants followed the instruction to perform as “quickly as possible” more vigorously than the instruction “as accurately as possible”. Our results rather showed a flat line close to zero (cf. Vanhove, 2014). La Brozzi (2012) did also not find a significant influence of inhibitory control on processing strategies measured by the Simon task.

Figure 7 shows the linear model for each cognitive variable and age. Crystallized intelligence (Gc) correlated strongly and significantly with age<sup>23</sup> ( $r(152) = .65, p < .001$ ). The negative correlation of fluid intelligence (Gf) and age was also significant ( $r(152) = .31, p < .001$ ). We did not find any development across the lifespan for working memory (WM) ( $r(152) = .03, p > .05$ ). Number of foreign languages (#L2s) was also significantly positively correlated with age ( $p < .001$ ).

Next, we examined whether responses to individual syllable conditions in terms of mean transformed hit rate correlated with the performance on the cognitive variables.

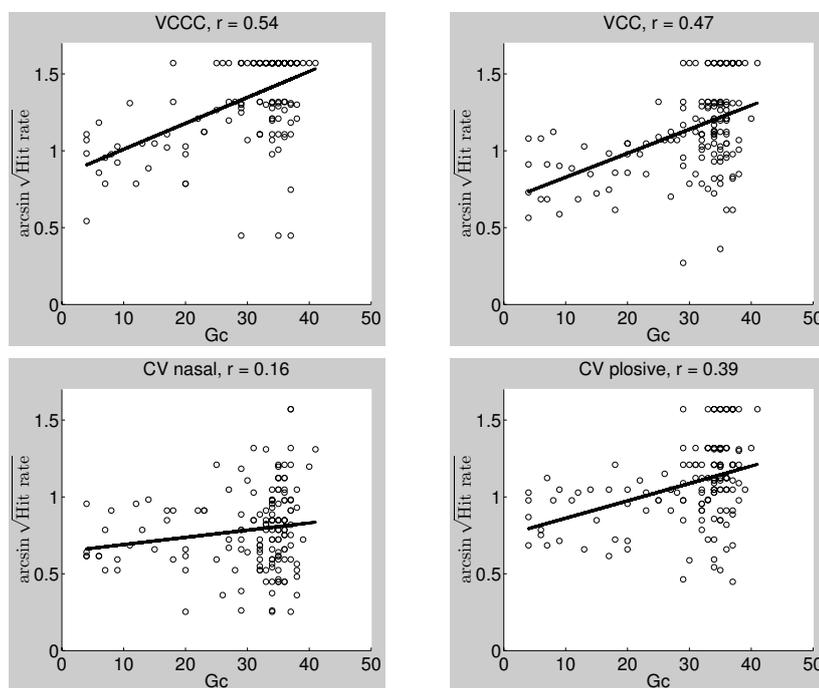


Figure 8: Linear correlation between crystallized intelligence and hit rate for each experimental condition,  $n = 152$ .

Figure 8 shows the correlation of crystallized intelligence (Gc) and hit rate for each syllable condition. Gc correlated very strongly with hit rate for VCCC ( $r(152) = .54, p < .001$ ), VCC ( $r(152) = .47, p < .001$ ) and CV\_plosive ( $r(152) = .39, p < .001$ ) and weakly for CV\_nasal ( $r(152) = .16, p < .05$ ).

<sup>23</sup> We are aware that the correlation of Gc and Gf along the lifespan would be more of a curvilinear correlation (cf. e.g. Baltes, 1987) and the current study is not meant to question this relationship. However, we are more interested in the general trend, which is why the linear relationship suffices to illustrate this point.

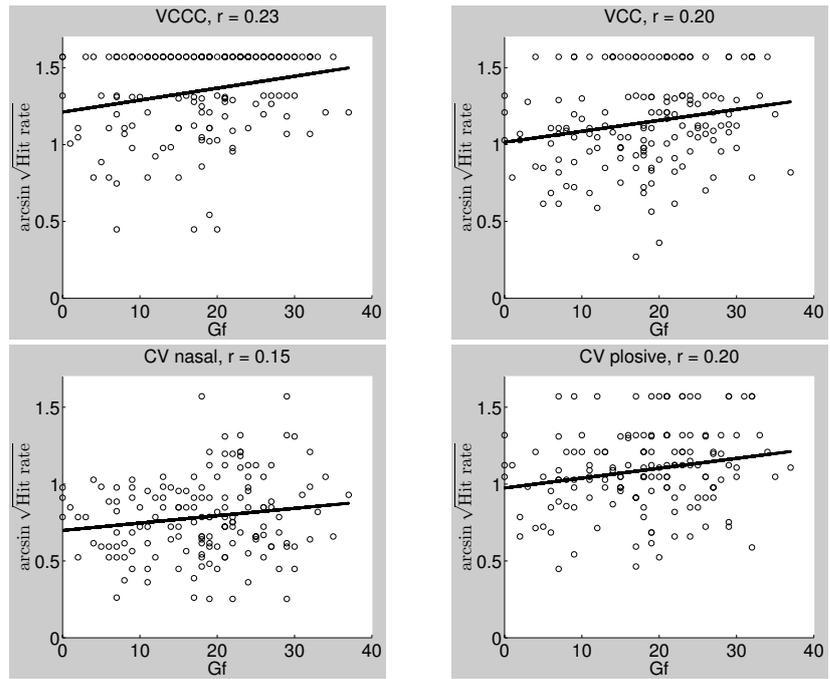


Figure 9: linear model for fluid intelligence and probability of correct rejection in each experimental condition,  $n = 152$ .

Figure 9 shows the correlation of fluid intelligence (Gf) and hit rate for each syllable condition. The correlation was moderate for VCCC ( $r(152) = .23, p < .01$ ), VCC ( $r(152) = .20, p < .01$ ) and CV\_plosive ( $r(152) = .20, p < .01$ ) and weak for CV\_nasal ( $r(152) = .15, p < .05$ ).

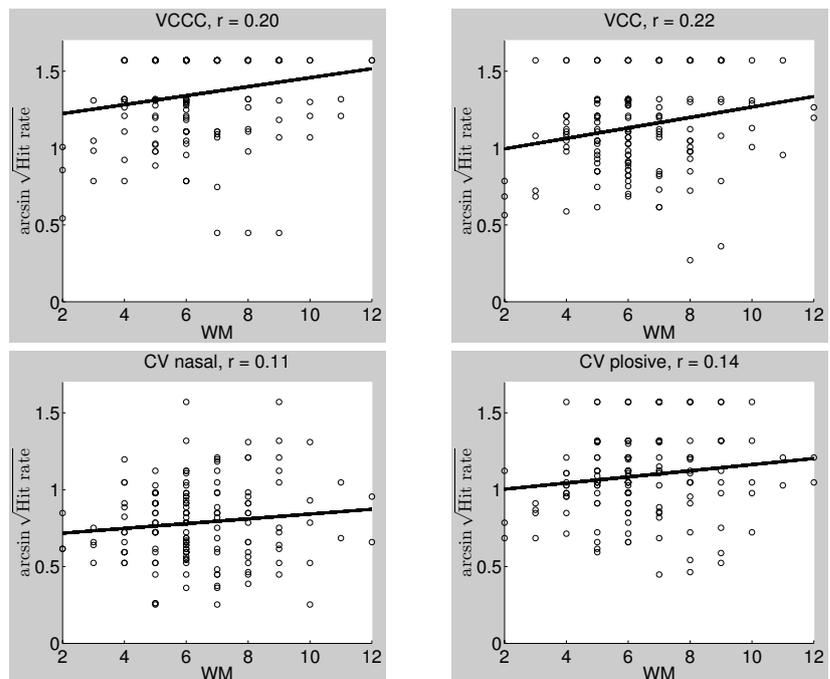


Figure 10: Linear correlation between working memory and hit rate for each experimental condition,  $n = 152$ .

Figure 10 shows the correlation of working memory (WM) and hit rate for each syllable condition. The correlation was moderate for VCCC ( $r(152) = .20, p < .01$ ) and VCC ( $r(152) = .22, p < .01$ ). The correlation for condition CV\_plosive and the CV\_nasal syllables did not reach significance (CV\_plosive:  $r(152) = .11, p > .05$ ; CV\_nasal:  $r(152) = .14, p > .05$ ).

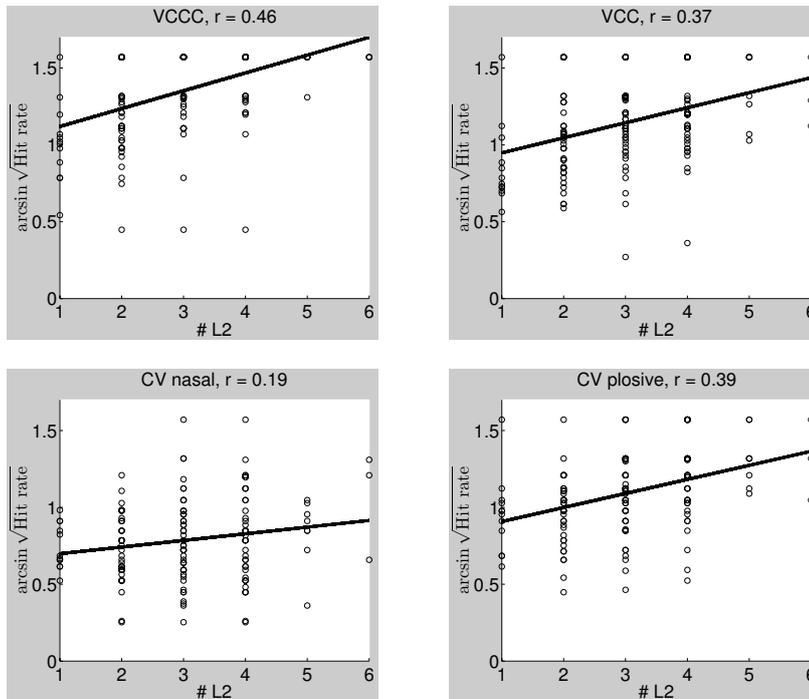


Figure 11: Linear correlation between number of foreign languages and hit rate for each experimental condition,  $n = 152$ .

Figure 11 shows the correlation of the number of foreign languages (#L2; Standard High German not included) and hit rate for each syllable condition. The correlation was strong for VCCC ( $r(152) = .46, p < .001$ ), VCC ( $r(152) = .37, p < .001$ ) and CV\_plosive ( $r(152) = .39, p < .001$ ), and weak for CV\_nasal ( $r(152) = .19, p < .05$ ).

### 3.4.7 Summary

We found that adults and children learned to distinguish sound regularities after seven minutes of exposure. In terms of accuracy, participants (as a group) could classify some syllable conditions more easily than others ( $VCCC > VCC = CV\_plosive > CV\_nasal$ ; where ' $>$ ' means higher accuracy and '=' no significant difference). We observed that accuracy increased across the lifespan in almost all syllable conditions ( $VCCC$ ,  $VCC$  and  $CV\_plosive$ ). Whenever the participants gave a correct response, they also tended to answer faster: the reaction time data nearly mirrored the results on accuracy ( $VCCC > VCC > CV\_plosive = CV\_nasal$ ; where ' $>$ ' means higher reaction time and '=' no significant difference), and increased in all conditions except  $VCCC$ .

Turning to the cognitive variables, findings both met and did not meet expectations. Crystallized and fluid intelligence and the number of foreign languages correlated with age, as expected. Unexpectedly, however, we found no correlation across the lifespan for working memory scores. Also against our expectations, crystallized intelligence strongly correlated with syllable conditions  $VCCC$ ,  $VCC$  and  $CV\_plosive$  and correlated weakly with  $CV\_nasal$  (the order of strength of correlations ' $VCCC > VCC > CV\_plosive > CV\_nasal$ ' was as expected). We expected the correlations of accuracy to be stronger with fluid than with crystallized intelligence. However, correlations with fluid intelligence and syllable-conditions were only weak. For working memory, we expected stronger correlations with accuracy in the four syllable conditions, yet correlations only reached significance in conditions  $VCCC$  and  $VCC$ . Surprisingly, the correlations with the number of foreign languages were very strong in syllable conditions  $VCCC$ ,  $VCC$  and  $CV\_plosive$  and there was a weak correlation with syllable condition  $CV\_nasal$ .

### **3.5 Discussion**

Study 1 asked three questions. The first question was how quickly learners can learn to distinguish words from non-words in an unknown, natural language that has not been pedagogically simplified for them and whether they can generalize knowledge that they acquired from the input to new stimuli. Second, we asked whether this capacity changed across the lifespan. Third, we examined whether cognitive variables such as crystallized and fluid intelligence, working memory and the number of foreign languages a person ‘knows’ affected the ability to distinguish ‘legal from illegal’ phonotactics in an unknown language. We found that participants could distinguish words from non-words and were able to generalize the newly acquired phonotactic knowledge to new words of the unknown language after only seven minutes of input. We also found that this capacity remained stable across the lifespan for the distinction of the most difficult syllables and even increased for the less difficult ones. The correlation with crystallized intelligence and the number of foreign languages and accurate responses was positive, while the correlation with fluid intelligence was negative, and the correlation with working memory did not change across the lifespan or for any syllable condition.

#### **3.5.1 ‘Learning’ after minimal exposure**

The results revealed implicit acquisition of Chinese phonotactics after seven minutes of exposure to audio-visual stimuli. The significant performance above chance level, both for words and non-words, showed that participants were able to generalize phonotactic knowledge that they had acquired through this brief exposure to new stimuli. It is particularly noteworthy that participants performed equally well on the CV\_plosives as on the VCC, and that they responded to illegal CV\_plosives significantly above chance. This suggests that participants derived phonotactic knowledge from the input, since the structure per se is possible in the participants’ native language, but the specific instantiation of CV\_plosive is only illegal in the target language. Moreover, the ability to correctly reject CV\_plosive syllables increased throughout the lifespan to a similar degree like in condition VCC.

The performance on the CV\_nasal syllables remained constant across the lifespan in terms of accuracy. As a group, participants were guessing on the CV\_nasal syllables. Given the performance at chance level in condition CV\_nasal, one interpretation along Carroll’s propositions in her AIT could be that participants failed to correctly

reject the CV\_nasal syllables because of an unsuccessful parsing of the input (cf. Carroll, 1999, 2000, 2002, 2004). According to the AIT, failures in L2-appropriate segmentation or acquisition of L2 knowledge occur when rules are inadequate or missing. This could result from incomprehensible or insufficient quantity of input. In consequence, their rules for categorization and reorganization of CV\_nasal syllables could not be revised and activated properly. Possibly, a higher quantity of input, for example double exposure to the WR, would have been sufficient for correct rejection of CV\_nasal syllables above chance (cf. Roberts et al., 2010).

The fact that reaction times in condition CV\_plosive did not differ significantly from reaction times in condition CV\_nasal was surprising. This could mean that participants still needed more ‘effort’ or attention to process CV\_nasal and CV\_plosive syllables (as compared to VCC syllables that were equally well rejected as CV\_plosive syllables) (cf. e.g. Ellis & Sagarra, 2010). Unsurprisingly, reaction times increased across the lifespan (cf. e.g. Baltes, 1987), however, not in the VCCC condition. The performance on the VCCC syllables remained constant across the lifespan in terms of reaction time. This could be explained by the near-ceiling effect of accuracy in condition VCCC, meaning that this syllable condition was very easy to reject and was easier to reject with increasing age.

The findings from this study are consistent with findings from Roberts et al. (2010) with regard to the accuracy results of the three- and two-consonant cluster syllables in the LDT. Unlike Roberts et al. (2010), we split the CVC syllables into CV\_nasal and CV\_plosive and could thereby show how performance differed depending on the illegal coda consonant. In Roberts et al.’s study (2010), performance was at chance for the CVC syllables. With regard to the number of L2s, our results differ from Gullberg et al.’s (2012) who did not find an effect in their model. Our study further differs from both Roberts et al. (2010) and Gullberg et al. (2012) with regard to the reaction time differences (between the conditions) that were non-existent in their studies. Importantly, beside the reaction time findings, we also contributed a lifespan perspective of the lexical decision performance after minimal exposure and correlation aspects with cognitive variables such as intelligence, working memory and the number of L2s (see below Section 3.5.2).

There is long-established knowledge of adults’ ability to use phonotactic information to decide whether a sound sequence is a ‘possible’ word or not (Greenberg & Jenkins, 1964). In a connectionist simulation model, Christiansen and colleagues

(1998) showed that phonotactic information on its own was enough to produce 47% accuracy in word segmentation, and that performance levels exceeded 70% accuracy when phonotactic information was combined with information of utterance boundary and relative stress. We measured our participants' lexical decision performance and thereby indirectly the segmentation ability. On average, our participants performed above 70% accuracy in all but the CV\_nasal conditions, even without prior training. The scores of all but the CV\_nasal condition lie within (and above) the typical accuracy-range for implicit learning (cf. e.g. DeKeyser, 2003, p. 319). We must therefore infer that participants were able to segment the complex, natural continuous speech in the WR and successfully generalize the implicitly learnt knowledge to new stimuli in the LDT. Swift learning after minimal exposure is further consistent with results in the artificial and statistical language learning literature (Friederici et al., 2002; Perruchet & Poulin-Charronnat, 2012; Saffran, Newport, et al., 1996; Saffran et al., 1997). Statistical language learning studies provide evidence that adults are sensitive to distributional cues and are able to learn incidentally as well as children (e.g. Saffran, Newport, et al., 1996; Saffran et al., 1997). In our study, adults even outperformed children in three out of four syllable conditions.

Three recent first exposure studies further support the view that adults are capable learners even if input is 'naturally' rich. Carroll and Widjaja's (2013) results suggest that there "are no absolute constraints on early stages" (p. 219) not only for representation (or perception), but also for production (examined in terms of mean accuracy scores only). In their study, participants were trained and tested on L2 (Indonesian) number constructions that largely differed from the respective L1 expressions. The results showed that some adult learners were able to acquire and internalize the constructions after only two training trials. Learners in our study did receive neither training nor feedback, which participants in Carroll and Widjaja's study received indirectly through the verification task. Rapid improvement and the ongoing ability to break into an unknown language after brief exposure was also demonstrated by Shoemaker and Rast (2013). They examined the learnability of phonological forms at the very initial stages of learning and the effect of utterance position and transparency of lexical items in classroom input. Their results suggest that learners rapidly developed sensitivity to the Polish phonetic system (in terms of mean accuracy scores only) and that as little as 1.5 hours per week of classroom instruction suffice for learners to begin to extract words from natural L2 speech. Compared to our controlled minimal input condition, Shoemaker and Rast's

exposure took place in a classroom environment in a total of 6.5 hours. Although input was also strictly controlled, the amount was substantially bigger and richer given the nature of the communication-based method that was used. It would have been interesting to see how learners in their study performed after smaller portions of input and not only at time 0 and after 6.5 hours. Bisson et al. (2013) reported very rapid incidental vocabulary learning through multi-modal exposure after only 320 trials, both in terms of accuracy and in terms of reaction time (in comparison: our LDT contained 256 test trials and 5 practice trials). The three studies are consistent with our findings of implicit learning in a naturally rich exposure condition. We cannot compare our results with regard to age span or cognitive variables, however, as Carroll and Widjaja (2013), Shoemaker and Rast (2013) and Bisson et al. (2013) only examined adults and did not administer any further tests.

### **3.5.2 Development across the lifespan**

Since late L2 learners with increased cognitive maturity were shown to overtake early learners (cf. e.g. Cenoz, 2002; Miralpeix, 2007; Muñoz, 2006), the discussion of the implicit learning development across the lifespan automatically runs alongside the discussion of the development of different cognitive variables that are influenced by age. Our results indicate that implicit learning scores are interrelated with crystallized and fluid intelligence scores, as well as with the knowledge of L2s, and partly also with WM. Moreover, our data suggests that age might be advantageous as a result of generally increased language experience, with L2s and with the L1, as well as an advanced cognitive capacity, wider attention spans and a greater abstraction capacity (cf. also Dimroth & Haberzettl, 2012; Muñoz, 2006; Snow & Hoefnagel-Höhle, 1978). Given the improved performance along the lifespan in the LDT, our results further support findings where adults have been shown to retain plasticity in the language system (e.g. Stein et al., 2006; Steinhauer et al., 2009) and the potential for acquiring high proficiency in a new language (e.g. Friederici et al., 2002; Montrul & Slabakova, 2003; Neufeld, 1977; Snow & Hoefnagel-Höhle, 1978; van Boxtel et al., 2005).

Our data regarding the cognitive variables only allowed us to interpret some of the results in terms of their predictive potential for correct rejection ability in the LDT. The scores of the crystallized and fluid intelligence task were consistent with the expected dispersion across the lifespan. The scores of the working memory task, however, showed no correlation with age across the lifespan. We expected an increase of WM for children and a decline of WM for adults (cf. e.g. Borella et al.,

2008; Jenkins et al., 1999; Park et al., 2002; Park & Payer, 2006). But we did expect the decline in WM scores for adults to be relatively small, because studies have found that the decrease in WM scores is mostly related to a decrease in processing speed (cf. e.g. Salthouse, 1996) and our WM task was not speeded. Recent literature has also made a distinction between visuo-spatial and verbal WM tasks and found different degrees of decline of WM scores with age. Hale et al. (2011), for example, found no evidence of age-related variance on verbal WM tasks, but age-related deficits on visuo-spatial processing in both speeded and unspeeded tasks. Alloway & Alloway (2013) also confirmed that WM skills across the lifespan are driven by the differences between the verbal and the spatial domain rather than by functional differences. Since the backward digit span task is a verbal WM task, our results are in line with these recent findings.

### ***Does crystallized intelligence influence accuracy?***

Our Gc-measure ('Wortschatztest') of crystallized intelligence tested participants' L1 mental lexicon size and was therefore mainly an indicator for general experience with the L1 (or the first L2 in the case of our participants, since the test was carried out in Standard High German). Results of the Gc correlated with age, as expected (cf. e.g. Baltes, 1987; Cattell, 1987; Horn & Cattell, 1967). The number of L2s also correlated with age. Since the highest correlations of Gc and number of L2s with accuracy were found in syllable conditions VCCC, VCC and CV\_plosive, we can presume that experience played a crucial role in foreign language learning in our LDT. The results could therefore be interpreted within the frame of advantages of a better-established L1 and/or the advantages of larger L1 mental lexicons (cf. e.g. Snow & Hoefnagel-Höhle, 1978). Best and colleagues (2001), for example, found that participants could better distinguish L2 sounds that were similar to L1 sounds. In a study by Frisch and colleagues (2001) that examined the role of lexical knowledge on the processing of non-words, participants with relatively larger mental lexicons were more likely to judge low probability non-words as well formed (cf. p. 164).

Because of the implicit nature of our experimental task, we did not expect that crystallized intelligence would correlate strongly with accuracy in all non-word conditions, especially not in conditions CV\_plosive and CV\_nasal. In studies by, for example, (McGeorge, Crawford, & Kelly, 1997), crystallized intelligence did not correlate with implicit learning (cf. also Reber et al., 1991). The correlations in conditions VCCC, VCC and CV\_plosive were much stronger than the correlations

with accuracy in these conditions with fluid intelligence. Given the high correlations of both crystallized intelligence (Gc) and the number of languages with accuracy in the four syllable conditions (VCCC > VCC > CV\_plosive > CV\_nasal), a possible explanation may be that no implicit learning was required to correctly reject VCCC, VCC and CV\_plosive syllables. If we only look at condition CV\_nasal, correlations with Gc and Gf with accuracy were almost identically weak. This would again conform to the hypothesis that Gc scores are not very predictive of implicit learning, since CV\_nasal is the condition that could not be correctly rejected with the aid of general knowledge, but was supposed to require implicit learning from the input. This interpretation would also be in line with findings by Roberts et al. (2010), where the no-input group could also reject conditions VCCC and VCC significantly above chance (cf. Discussion Study 2, Section 4.5.1).

### ***Does fluid intelligence influence accuracy?***

Our Gf-measure (Raven test) of fluid intelligence was a non-verbal perceptual reasoning task. We allowed participants to maximally work 60 minutes on the matrices. Kaufman et al. (2010) only gave a 45 minutes time allowance, but their participants were only 16-18 years old. We wanted to make sure that younger and older people alike had enough time to work on the task. Fluid intelligence decreased with increasing age as expected (cf. e.g. Baltes, 1987; Cattell, 1987; Horn & Cattell, 1967). The correlation with accuracy in the four syllable conditions was only moderate in conditions VCCC, VCC and CV\_plosive and weak in condition CV\_nasal. We would have expected higher correlations with fluid intelligence, because deductive reasoning is thought to be independent of acquired knowledge and crystallized intelligence (cf. Cattell, 1987). Fluid intelligence correlated with artificial grammar learning in the explicit, but not the implicit serial learning instruction condition in a study by Gebauer and Mackintosh (2007). The authors interpreted their results in support of Reber and colleagues' claim that implicit learning is independent of intelligence (1991). Robinson (2002), however, differentiated the interpretation of Reber et al.'s (1991) findings with his own work on individual differences in intelligence, aptitude, working memory, and incidental adult L2 learning. Robinson (2002) suggests that implicit but not incidental learning is related to intelligence, such that intelligence and implicit learning are correlated significantly negatively. The correlations in our study, however, point in the opposite direction. Better intelligence scores are positively related with higher accuracy in the LDT, especially in conditions VCCC, VCC and CV\_plosive. Fluid intelligence and accuracy in CV\_nasal correlated only weakly. This result is in line

with Misyak and Christiansen's (2012), who also did not find significant associations with fluid intelligence and incidental statistical learning in their study.

### ***Does working memory influence accuracy?***

We found no evidence that working memory correlated with age. A possible reason for this could be the nature of the WM task. Grégoire and Van der Linden (1997) did not find a significant effect of age on the difference between forward and backward digit span tasks, although age is supposed to have more influence on the backward digit span task, because the backward digit span task involves the capacity of inhibitory control. According to the Inhibition-Reduction theory, inhibition is responsible for age-related changes in cognitive performance (Hasher & Zacks, 1988; cf. also Persad, Abeles, Zacks, & Denburg, 2002). However, in a study by Borella and colleagues (2008), with participants between the ages of 20 to 86 years, inhibition accounted for a part of WM decline across the lifespan, but was not a crucial contributor to age-related changes of WM<sup>24</sup>. A possible alternative WM task could have been the computerized mean span metric of digit span discussed in Woods et al. (2010). This span task has an enhanced sensitivity of forward versus backward span comparisons. The Operation Span Task (cf. Turner & Engle, 1989) would have been a well-validated and probably even more reliable alternative WM task that we also considered. However, within our extensive test-battery, this task would have taken up too much time.

Even though working memory did not correlate with age, it did correlate with accuracy in syllable conditions VCCC and VCC, almost equally strongly as fluid intelligence did. This result is in line with Engle et al.'s (1999) finding that fluid intelligence and working memory show strong connections. We can, however, not infer that the function of WM is to act as a language-learning device from our results (Baddeley, Gathercole, & Papagno, 1998), because there was no significant correlation with WM and the two conditions (CV\_nasal and CV\_plosive) which assume that learning must take place in order to perform accurately. In studies such as Misyak and Christiansen (2012), for example, verbal WM was seen as an "index of processing skill for language comprehension and statistical learning" (p. 321). Similarly, L2 development occurred in those learners with the highest WM scores in Mackey and Sachs's study (2012).

---

<sup>24</sup> Since our results of executive control (measured in the Simon task) were not evaluated, our data does not allow us to confirm or contradict the inhibition-reduction theory (Hasher & Zacks, 1988).

Another possible explanation for our partial lack of correlation with WM could be related to the implicit nature of our task. In Robinson (2002), working memory did not predict successful incidental learning. Yet, Robinson (2002) also found significant negative correlations of intelligence scores and incidental learning, which we did not. Unsworth and Engle (2005) corroborated Robinson's finding by demonstrating that no WM differences emerged in incidental learning conditions. At the same time, they showed WM differences in implicit learning conditions in a serial reaction time task that was instructed explicitly. This is more in line with our findings.

Overall, our results speak against DeKeyser's claim that adults are no longer capable of implicit learning (DeKeyser, 2003). Since we observed an increase of the ability beyond the ages of early puberty, our results further contradict Janacek et al. (2012) who observed a rapid decrement of implicit abilities around the age of twelve. The results, therefore, are not in line with findings of a general decline in L2-learning ability and proponents of an early age of onset (AO) advantage (e.g. Abrahamsson & Hyltenstam, 2009; Hyltenstam & Abrahamsson, 2003). What is more, contrary to popular beliefs and to the literature dealing with age effects in acquisition, our findings suggest that this ability seems to remain stable across the lifespan. The so-called 'qualitative shift from implicit to explicit' (DeKeyser, 2012, p. 456) does not appear in our data, and may indeed cut across the statement that there is "little hard [empirical] evidence of learning without awareness" in general (DeKeyser, 2003, p. 317) and especially for the implicit learning of abstract structures by adults (cf. p. 321).

### *Is there a multilingual advantage?*

In the post-experimental debriefing, most participants said that they distinguished between 'English' words and Chinese words in the LDT when we asked them if they made use of any meta-linguistic strategies (i.e. "Did you use a strategy to decide whether a word was Chinese or not?"). This implies that participants did incidentally pay attention to the phonotactics and implicitly made inferences of 'what is Chinese' and what is not. The correlation of the number of L2s with accuracy in the LDT further points towards an advantage of knowing more L2s when starting to break into another language at first contact.

### 3.5.3 Conclusion

In conclusion, we found that people across the lifespan can learn to distinguish sound regularities in natural language input after seven minutes of exposure. What is more, they are able to generalize the acquired abstract information to new stimuli of the unknown language. Surprisingly, this ability seems to improve with increasing age. However, it is important to specify that what improved with age was the ability to reject L1-sounding words, like three-consonant clusters, as being Chinese. The ability to correctly reject CVC non-words only partly improved with age. CV\_plosive syllables improved across the lifespan to a similar degree as VCC syllables. It is particularly remarkable that illegal CV\_plosives were responded to significantly above chance. This suggests that participants derived phonotactic knowledge from the input, since the structure per se is possible in the participants' native language and in the target language, but the specific instantiation of CV\_plosive is only illegal in the target language. The ability to reject CV\_nasal syllables did not improve with increasing age. Importantly, however, that ability also did not decline, but remained constant across age. The explanation for these findings may be found in aspects of higher crystallized intelligence and/or an increase in number of L2s across the lifespan, but that remains to be studied further. At any rate, the two findings together still speak against a simplistic critical period account of the perception and generalization of newly acquired phonotactic knowledge to non-native language input.

Since all our participants were exposed to the same input condition and we did not test a control group that did not receive the input (WR), we cannot conclude with certainty that the above chance performance in Study 1 originated from implicit learning during the minimal exposure. Furthermore, we compared the ability to distinguish sound regularities across the lifespan, but did not specifically compare different age groups. To corroborate age effects in the different syllable conditions, it would be fruitful to compare at least two age groups. Moreover, testing a control group would have allowed us to tie effects back to the exposure condition. In case of a learning effect from input, we could then investigate possible consolidation effects to test whether this learning was only temporary or became embedded in memory.

## **4 Influence of the input (Study 2)**

### **4.1 Introduction and Background**

Study 1 explored the development of the ability to extract phonotactic regularities from continuous natural speech at first exposure across the lifespan. All participants in Study 1 first listened to the Weather Report (WR) before conducting the Lexical Decision Task (LDT). The above-chance performance on the LDT, especially on critical CV\_plosive non-words, was interpreted as learning effects from the seven minutes of speech input. However, Study 1 did not compare the performance of participants to a control group that has not received any input. Under these circumstances, we cannot exclude the possibility that participants simply inferred which words were Chinese or not without having learnt anything from the exposure. In Study 2, we therefore focus on the following three main questions: First, is there evidence of real learning from input or is general inferencing enough? Second, do children differ from adults in this regard? Third, is there evidence of real learning with consolidation of phonotactic knowledge after one week?

Study 2 is partly a replication of Roberts et al.'s (2010) study. They used the same experimental paradigm and also compared a control group that did not receive any input to an exposure group. Furthermore, they examined whether there was a significant increase in learning after double exposure to the Weather Report. However, they did not test a child control group and did not correlate performance on the LDT with any cognitive variables. They did also not test participants a second time on the LDT after a consolidation period.

#### **4.1.1 From input to consolidation**

Given Corder's (1967) differentiation of input and intake, Study 2 examined whether possible learning effects could be traced back to the input we had provided in the Weather Report, or whether previous intake experiences or knowledge suffice to make lexical decisions significantly above chance level. Memory and sleep consolidation research has also shown how training effects of novel phonological contrasts and words can be enhanced through retention intervals that contain sleep periods (cf. e.g. Davis et al., 2009; Fenn et al., 2003; Tamminen et al., 2010). Sleep even seems to evoke memory consolidation and lexical integration processes that are

similar to the effects of training tasks (Lindsay & Gaskell, 2009). Behavioural and neuroimaging research has further shown how different aspects in the process of learning new words became established over different time scales (cf. Lindsay & Gaskell, 2010). To our knowledge, however, there have not been many studies comparing adults' and children's performance on lexical learning after (sleep) consolidation periods without prior training, especially not using natural language stimuli.

Declarative memory consolidation after sleep was demonstrated in children by Backhaus (2008). When compared for implicit sequence learning, adults outperformed children on a serial reaction time task (cf. Thomas et al., 2004). Fischer, Wilhelm and Born (2007) compared the effect of sleep on implicit memory formation in 7-11-year-old children and 20-30-year-old adults with a serial reaction time task. They found a striking contrast between children and adults insofar, as differences in reaction time between grammatical and non-grammatical trials decreased in children, but increased in adults after sleep. Thus a gain of knowledge through sleep in adults contrasted a sleep-dependent deterioration in children in measures of implicit sequence knowledge. This contrast did not arise in the wake retention period. For implicit memory consolidation, the authors therefore concluded that the functional role of sleep is age-dependent. Wilhelm, Diekelmann and Born (2008) confirmed this result comparing children and adults on declarative and procedural memories after daytime retention and nocturnal sleep. While sleep similarly affected declarative memories in children and adults, adults improved comparably more on procedural skills than children after sleep than after wake (cf. also review on 'the whats and whens of sleep-dependent memory consolidation' by Diekelmann, Wilhelm, & Born, 2009).

In sum, age effects are currently a hotly debated topic not only in the language acquisition literature, but also in research on consolidation effects. Our study contributes new information to both discussions of implicit learning skills in children and adults in the following ways. Firstly, we compare the effect of input on children and adults between an experimental and a control group. Secondly, we compare incidental implicit learning effects of natural language stimuli in children and adults. Thirdly, we compare these effects after a consolidation time of one week.

#### **4.1.2 Research questions**

- 1) Is there evidence of real learning from input or is general inferencing enough?
- 2) Do children differ from adults in this regard?
- 3) Is there evidence of real learning with consolidation of phonotactic knowledge after one week?

## 4.2 Methods

### 4.2.1 Participants

For Study 2, we recruited 80 participants (39 women, 41 men) in two different age groups (out of 120 screened)<sup>25</sup>: children between 10 to 11 years, and adults between 30 to 40 years of age. Participants were randomly assigned to either the control or the experimental group, each consisting of 40 persons. Of these, 68 (32 women and 36 men) were retained for analysis, 32 in the experimental group and 36 in the control group. Participants in the experimental group agreed to participate twice: once at time 1 (T1) and a second time exactly one week later, at time 2 (T2). The control group only participated once. All participants provided written consent, which, in the case of children, was provided by their parents.

Participants were recruited using the same selection criteria as in Study 1 (cf. Section 3.2.1). Recruitment was done via multiple channels: 1) email lists, 2) an (online) bulletin board of city council, 3) announcements displayed around the university campus, and 4) by word of mouth.

Participants who met the screening criteria were asked to fill out the second language background questionnaire prior to the actual experiment. Due to time constraints, we decided to shorten the extensive (second) language background questionnaires from Study 1. The briefer version was sent out by email prior to the first appointment and returned to us either via email or printed out. Participants in the child groups filled out the questionnaire at the end of the experiment to allow time for possible questions. The language background questionnaire lasted about 10 minutes. In the first question, participants self-assessed their listening and reading capacity of all the languages and dialects. Each language level of competence of the Common European Framework of Reference for Languages (CEFR)<sup>26</sup> was described by two brief sentences only, following observations in Study 1 of participants' inability to cope with the detailed descriptions of the different levels. In a second

---

<sup>25</sup> In total, 120 people had profiles suitable for participation in our experiments; however, many could not participate on two different dates (experimental group) and/or assist during pertinent time slots.

<sup>26</sup> A1 (ab initio), A2, B1, B2, C1, C2 (proficient)

step, participants were asked to indicate in what context they acquired their L2s<sup>27</sup> and dialects (except for Swiss-German) — through school, a language course, an exchange abroad, in direct contact with native speakers, or through media. Table 8 summarises participants’ knowledge of other languages. The third question asked how frequently they used each language on a scale of 5 with ‘1’ indicating almost never, and ‘5’ very frequently. Finally, they were asked at what age they first used the languages, and to indicate their highest educational degree. As in Study 1, the information from the second questionnaire will not be evaluated further in our study.

Table 8: Participant characteristics in terms of the distribution of gender, age, average number of foreign languages (Mean #L2s), and which foreign languages (L2s) were spoken by how many participants in each group,  $n = 68$ .

Age	Total	Male	Female	Mean age	Mean # L2s	L2s
<i>Experimental</i>						
10-11	14	6	8	11	2.36	English (14), French (14), Italian (1), Spanish (2), Albanian (2), Serbian (1)
30-40	18	8	10	33	2.61	English (18), French (18), Italian (4), Spanish (5), Portuguese (1), Ivrit (1)
<i>Control</i>						
10-11	20	14	6	10	2.4	English (20), French (14), Italian (3), Arabic (1), Dutch (1), Hungarian (1), Rhaeto-Romance (1), Serbian (1)
30-40	16	8	8	32	2.81	English (16), French (16), Italian (8), Spanish (5)
Total	68	36	32	21.5	2.55	English (68), French (68), Italian (16), Spanish (13), Albanian (2), Serbian (2), Dutch (1), Hungarian (1), Ivrit (1), Portuguese (1), Rhaeto-Romance (1)

Children in the experimental group were paid 20 CHF, children in the control group 15 CHF, adults in the control group 40 CHF, and adults in the experimental group 50 CHF for their participation in the experiment.

#### 4.2.2 Experimental Materials

##### *Exposure: Weather Report*

The same Weather Report (WR) was used as in Study 1 (cf. Section 3.2.2).

<sup>27</sup> As in Study 1, we will refer to any additional languages beside the first language as L2s from now on.

### ***Experimental testing: Mandarin Chinese Lexical Decision Task***

The same Lexical Decision Task (LDT) was used as in Study 1 (using the same headphones, the same button-box and the same software, cf. Section 3.2.2).

### **4.2.3 Cognitive Test: additional Material**

#### ***Working memory (WM)***

The same Working memory task was used as in Study 1 (cf. Section 3.2.3).

### **4.2.4 Procedure**

Four groups were tested: one experimental child group, one control child group, one experimental adult group, and one control adult group. Adult participants were mainly tested at the University of Zurich. A quiet room was provided where disturbances could be kept minimal. Participants were always asked to switch off their mobile phones. Sometimes, more than one participant was tested at the same time: up to four participants could be comfortably seated in the same room. In cases of multiple testing, participants were instructed together at the beginning or separately outside the room and asked to whisper if they had additional questions during the experiment. Children participants were mainly tested at three different schools in the region of Zurich city. In each of the schools, two quiet rooms were provided. Children were closely watched to keep mutual eye contact to a minimal and reminded of the proper posture and serenity, if necessary.

There were two experimental tasks: the Chinese Lexical Decision Task and the backward digit span task. The order of the two experimental tasks was randomized throughout the four groups. The experimental groups listened to the Weather Report at time 1 and completed the Lexical Decision Task (LDT) immediately thereafter (as in Study 1, cf. Section 3.2.4). At the second testing, the experimental groups were only asked to do the LDT again – without listening to the WR and without re-doing the backward digit span task. Participants in the control groups did not watch the WR and only did the LDT and the backward digit span task.

During multiple testing, there were at least two examiners present to ensure prompt supervision and to avoid eye contact and mutual distraction between the participants.

## 4.3 Data Analysis

### 4.3.1 Data cleaning

Initial data inspection involved the identification of participant outliers. We excluded one participant due to technical problems, two due to biased response patterns, four because of due to missing data on the cognitive variable, and five participants due to zero correct responses in any syllable condition of our experimental task (at any measurement time<sup>28</sup>). This left us with 68 out of 80 participants, which account for 85% of the collected data.

Four more steps of data cleaning and transformation were done, identical to Study 1 (cf. Section 3.3.1):

- Reaction times smaller than 100 ms and bigger than 2000 ms as well as trials with an onset delay bigger than 250 ms were removed as these answers constitute technical errors. This step affected about 2% of trials.
- Reaction times were then log-transformed since their empirical distribution was slightly left-sided (see Figure 12).
- Winzorizing reaction times above or below two standard deviations of the group mean of the specific conditions. This procedure affected on average 3.6% of the data per condition, across all groups.
- Accuracy was transformed with arcsine-square-root to restore the assumption of normality.

No further data cleaning was undertaken (and importantly, no winzorizing or trimming of the accuracy data).

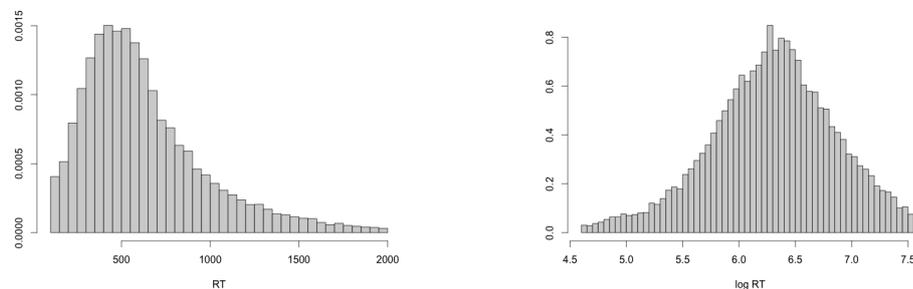


Figure 12: left panel: raw reaction times across all subjects, across all conditions after removal of responses  $<100\text{ms}$  and  $>2000\text{ms}$ ; right panel: log-transformed reaction times.

---

<sup>28</sup> All five participants had zero accuracy in one of the syllable conditions, either at time 1 or at time 2, which meant that these participants were excluded altogether.

### 4.3.2 Statistical analyses

The results from the Lexical Decision Task were analyzed in terms of accuracy and reaction times. We started by comparing untransformed mean hit rate and reaction times. Next, hit rates were arcsine-square-root-transformed, reaction times were log-transformed, and the four non-word conditions VCCC, VCC, CV\_plosive and CV\_nasal were compared across groups, as in Study 1 (cf. Section 3.3.2). The real words served as filler items. Of the 128 filler items, 64 have appeared in the Weather Report and were not investigated further. The 64 filler items that were heard for the first time in the LDT served as a reference category to compare the overall accuracy on words and non-words<sup>29</sup> (non-words were all heard for the first time in the LDT). The only analysis carried out with the filler items was, first, an omnibus ANOVA to control that fillers were answered equally (correctly) as the experimental items, across experimental and control groups and across children and adults. Second, a *t*-test for each condition examined whether the mean accuracy differed significantly from chance level. This *t*-test was carried out once for the filler items and once for the four non-word conditions, across all group means taken together, both in terms of mean accuracy (in arcsine-square-root) and in terms of mean (log-transformed) reaction times. We considered only reaction times of the trials with correct answers.

Next, an omnibus ANOVA examined whether the non-word means differed significantly between conditions, groups and ages. As in Study 1, post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007) (cf. Section 3.3.2 for description of Tukey). The strength of the post-hoc multiple comparisons was reported with Cohen's *d* (Cohen, 1988).

In a next step, a repeated-measures ANOVA tested how testing-time (time 1 versus time 2) and age affected the hit rate and the mean reaction times given a syllable condition, and whether within-subject comparisons differed for time 1 versus time 2 in both the children and the adult group, for each syllable condition.

In a last step, we investigated the correlations of the cognitive variables with age and with accuracy and reaction times in both groups (experimental and control) in each syllable condition using Pearson Correlations. To simplify the measure of 'multilingualism', we only tested whether the number of foreign languages

---

<sup>29</sup> Since we only analyzed non-word conditions VCCC, VCC, CV\_nasal and CV\_plosive, the number of total non-words also added up to 64 non-words.

(‘Number of L2s’) contributed to the prediction of our dependent variables. Working memory scores from the digit span task were treated as numbers and we only considered the total number of correct responses. We did not analyze the digit span.

#### **4.3.3 Hypothesis**

- 1) The experimental groups perform significantly better than the control groups on the Lexical Decision Task; both
  - a. in terms of accuracy, and
  - b. in terms of reactions time.
  - c. But NOT in terms of working memory and the number of L2s.
  
- 2) Adults outperform children.
  - a. It is easier to classify VCCC syllables as non-words than VCC syllables.
  - b. It is easier to correctly classify VCC syllables as non-words than CVC syllables.
  - c. It is easier to correctly classify CV\_plosive syllables as non-words than CV\_nasal syllables.
  - d. 2a., 2b., and 2c. concern accuracy, but NOT reaction times.
  
- 3) The performance on the LDT stays stable or improves after one week.

## 4.4 Results

As in Study 1 (cf. Section 3.4), we report the results on the same four non-word syllable conditions, namely the two ‘control’ conditions VCCC and VCC, and the two ‘critical’ conditions CV\_nasal and CV\_plosive (= 64 items; cf. Table 9 as a reminder of the four conditions). The results on the real words serve as reference values for the results on the non-words. We therefore only refer to the real words that have appeared for the first time in the LDT (= 64 items). If we refer to the non-words as compared to the real words, we always mean the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) taken together.

In the first section, we will report results across and between the four tested groups (Experimental Children, Experimental Adults, Control Children, Control Adults) at time 1. In a second section, time 1 (T1) and time 2 (T2) are compared between Experimental Children and Experimental Adults.

Table 9: The four non-word conditions in the LDT. The illegal consonants are underlined.

	Items	Consonant cluster	Example	Items	Consonant cluster	Example
‘Control’	16	V <u>CCC</u>	<i>alst</i>	16	V <u>CC</u>	<i>ans</i>
‘Critical’	16	C <u>V</u> (illegal nasal)	<i>gam</i>	16	C <u>V</u> (illegal plosive)	<i>mat</i>

### 4.4.1 Mean Accuracy and Reaction times per syllable condition – T1 & T2

Table 10: Mean and standard deviation (in brackets) of untransformed hit rates on the Lexical Decision Task per non-word syllable condition and real words.

	Control groups		Experimental groups T1		Experimental groups T2	
	Children	Adults	Children	Adults	Children	Adults
VCCC	.768 (.147)	.970 (.054)	.807 (.120)	.946 (.005)	.639 (.285)	.992 (.034)
VCC	.619 (.184)	.798 (.128)	.569 (.174)	.794 (.133)	.557 (.285)	.873 (.129)
CV_nasal	.497 (.234)	.462 (.210)	.465 (.193)	.619 (.199)	.400 (.157)	.696 (.215)
CV_plosive	.541 (.178)	.757 (.090)	.467 (.149)	.865 (.133)	.458 (.191)	.858 (.182)
Fillers	.617 (.122)	.710 (.117)	.681 (.096)	.663 (.116)	.626 (.102)	.620 (.131)

Table 11: Mean and standard deviation (in brackets) of untransformed reaction times on the Lexical Decision Task per non-word syllable condition and real words.

	Control groups		Experimental groups T1		Experimental groups T2	
	Children	Adults	Children	Adults	Children	Adults
VCCC	554 (119)	542 (231)	536 (127)	443 (118)	508 (114)	401 (106)
VCC	626 (190)	623 (206)	588 (148)	531 (122)	549 (242)	484 (122)
CV_nasal	732 (218)	785 (190)	614 (161)	727 (207)	615 (205)	687 (195)
CV_plosive	763 (233)	803 (262)	641 (235)	660 (135)	620 (165)	641 (161)
Fillers	714 (161)	724 (200)	670 (121)	629 (142)	675 (136)	630 (143)

We started by computing the mean hit rate (i.e. accuracy; correct rejections for non-words and correct hits for fillers) and the mean reaction times to correct trials in the LDT across all participants and ages. Table 10 summarises the untransformed mean proportions and

Table 11 summarises the untransformed reaction times per condition.

#### 4.4.2 Across groups – Comparison of syllable conditions (T1)

The mean proportions of accuracy were transformed by arcsine-square-root and the reaction times by natural logarithm. Accuracy at chance level for untransformed accuracy scores equals .50, but equals .79 with arcsine-square-root. Next, we examined whether the mean response accuracies of the non-words (VCCC, VCC, CV\_nasal and CV\_plosive taken together) differed across the four groups. Figure 13 summarises the arcsine-square-root-transformed mean hit rates and mean log-transformed reaction times by condition collapsed across all participants at time 1 only.

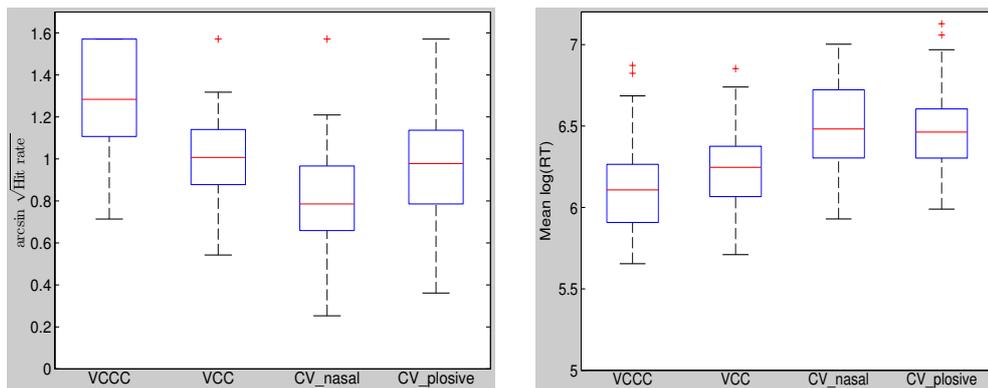


Figure 13: Boxplots comparing (left) mean transformed hit rates and (right) mean transformed reaction times, averaged across all four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1,  $n = 68$ ), for all four conditions (VCCC, VCC, CV\_nasal, CV\_plosive).

As in Study 1 (cf. Section 3.4.2), Figure 13 (left) suggests that phonotactically illegal syllables with three-consonant-clusters (VCCC) across all groups were easier to identify than illegal two-consonant-clusters (VCC). Two-consonant-clusters (VCC) were easier to identify than CV\_nasal syllables, and CV\_plosive syllables were easier to identify than CV\_nasal syllables.

Analogous to Study 1, Figure 13 (right) also suggests that reaction times mirrored the accuracy scores. Responses were fastest to phonotactically illegal syllables with the three-consonant-cluster (VCCC), followed by illegal two-consonant-cluster-syllables (VCC), and CV\_plosive and CV\_nasal syllables.

### ***Accuracy across all groups at T1***

After Levene's test for equal variances was confirmed ( $F = .97, p = .41$ ), an ANOVA<sup>30</sup> with Syllable Condition (VCCC, VCC, CV\_nasal and CV\_plosive) as the independent and Accuracy as the dependent variable was carried out to test whether there was a significant difference in mean hit rate for the four conditions, across the four groups. There was a significant main effect for Syllable Condition ( $F(3, 268) = 44.3, p = .00001$ ). In other words, the response accuracy varied for the different syllable conditions, in like manner of Study 1.

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007).

Table 12: Tukey's honestly significant difference criterion for mean transformed hit rate, between the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1).

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
VCCC - VCC	0.272	0.163	0.380	1.129	1
VCC - CV_nasal	0.210	0.102	0.319	0.881	1
VCC - CV_plosive	0.037	-0.072	0.145	0.149	0
CV_nasal - CV_plosive	-0.174	-0.283	-0.065	0.686	1
VCCC - CV_nasal	0.482	0.373	0.591	1.928	1
VCCC - CV_plosive	0.308	0.199	0.417	1.207	1

Tukey's HSD criterion showed that the mean hit rate for each syllable condition was significantly different from every other syllable condition, with the exception of the comparison between CV\_plosive and VCC (cf. Table 12). We observed the same relationship in Study 1.

### ***Reaction times across all groups at T1***

After Levene's test for equal variances was confirmed ( $F = .93, p = .43$ ), an ANOVA with Syllable Condition as the independent and Reaction Time as the dependent variable was carried out to test whether there was a significant difference in mean transformed reaction times for the four conditions, across the four groups. There was a significant main effect for Syllable Condition ( $F(3, 268) = 34.6, p = .00001$ ) parallel to results obtained in Study 1. In other words, the reaction times varied for the different syllable conditions.

---

<sup>30</sup> For all reported ANOVAs, alpha was set at 0.05. For ANOVA-tables cf. Appendix C.

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007).

Table 13: Tukey's honestly significant difference criterion for mean transformed reaction times in between the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1).

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
VCCC - VCC	-0.137	-0.250	-0.024	-0.562	1
VCC - CV_nasal	-0.241	-0.353	-0.128	-0.949	1
VCC - CV_plosive	-0.223	-0.336	-0.110	-0.879	1
CV_plosive - CV_nasal	0.018	-0.095	0.130	-0.067	0
VCCC - CV_nasal	-0.378	-0.490	-0.265	-1.469	1
VCCC - CV_plosive	-0.360	-0.472	-0.247	-1.400	1

Tukey's HSD criterion showed that each syllable condition was significantly different from every other syllable condition in terms of mean reaction times, except for CV\_plosive from CV\_nasal (cf. Table 13), which we also observed in Study 1.

#### 4.4.3 Comparison to chance level across groups at T1

In a third instance, we examined whether the mean hit rate differed significantly from chance. Bonferroni corrected right-tailed Student's *t*-tests revealed that the accuracy scores for three of the syllable conditions were significantly different from chance (VCCC  $t(67) = 16.36, p < .001$ ; VCC  $t(67) = 8.20, p < .001$ ; and CV\_plosive  $t(67) = 6.10, p < .001$ ). In contrast, responses to CV\_nasal syllables were at chance ( $t(67) = .58, p > .05$ ). The results suggest that participants overall were able to correctly reject the consonant-cluster syllables as not being Chinese, and that they were also able to identify the CV\_plosives as not being Chinese. As a group, they were randomly guessing on the CV\_nasal syllables. Note that illegal CV\_plosive syllables were responded to significantly above chance, as in Study 1 (cf. Section 3.4.3). The transformed hit rate for the four non-words taken together was significantly above chance ( $t(67) = 45.52, p < .001$ ), as well as the transformed hit rate for the real words that appeared for the first time in the LDT ( $t(67) = 11.12, p < .001$ ). These results are consistent with results in Study 1.

#### 4.4.4 Across groups – Words compared to non-words (T1)

##### *Accuracy – words-non-words; across groups, T1*

Next, we examined whether the mean transformed hit rate of the real words differed significantly from the non-words.

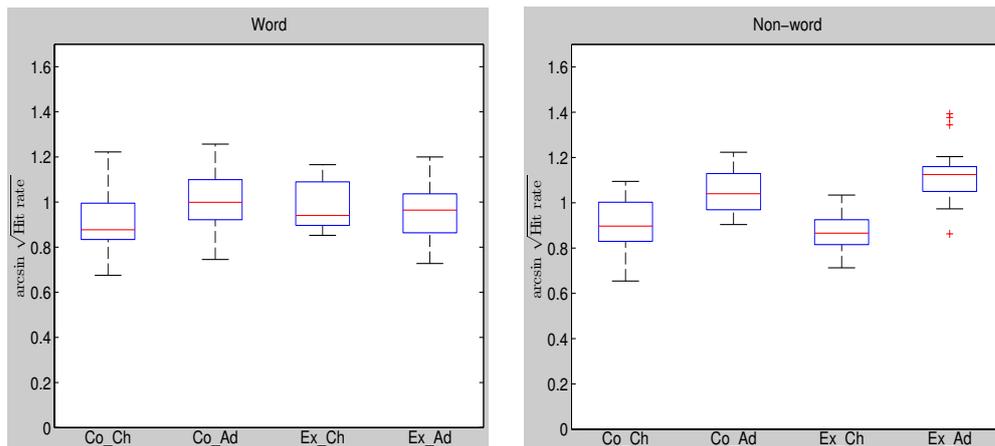


Figure 14: Boxplots to compare mean transformed hit rate for words and non-words, for all four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1;  $n = 68$ ). Transformed chance level equals .79.

Figure 14 shows the comparison of the four groups for words versus non-words. In both the Control and the Experimental group, adults showed higher hit rates than children, especially in the non-word conditions. An omnibus ANOVA<sup>31</sup> was conducted, including Group (control vs. experimental), Age (adult vs. child) and Word Type (word vs. non-word) as independent variables and Accuracy as dependent variable, after Levene's test for equal variances was confirmed ( $F = .349$ ,  $p = .030$ ). The ANOVA revealed a main effect of Age ( $F(1, 129) = 33.51$ ,  $p = .00001$ ), and a significant interaction of Age and Word Type ( $F(1, 129) = 13.94$ ,  $p = .0003$ ). There was no main effect for Group ( $F(1, 129) = .53$ ,  $p = .47$ ).

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007)<sup>32</sup>.

Table 14 summarises the results of the Tukey's test. The analyses showed that children in both the Control and the Experimental groups treated words and non-words equally, on average, but adults in both the Control and the Experimental groups, performed better on the non-words than children, and experimental adults' mean transformed hit rate was significantly higher on the non-words than on the words (which can be explained by adults' near-ceiling performance in condition VCCC, see below Section 4.4.5).

<sup>31</sup> For all reported ANOVAs, alpha was set at 0.05.

<sup>32</sup> Cf. Method Summary section of Study 1 for description of Tukey's test.

Table 14: Tukey's honestly significant difference criterion for mean transformed hit rate in the comparisons of words (W) and non-words (NW), across the four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1).

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
<i>Within Ex_Ch:</i>					
NW - W	-0.042	-0.162	0.078	-1.073	0
<i>Within Ex_Ad:</i>					
NW - W	0.119	0.009	0.008	1.315	1
<i>Between Ex_Ch and Ex_Ad:</i>					
W (Ch) - W (Ad)	-0.042	-0.158	0.074	0.164	0
NW (Ch) - NW (Ad)	-0.204	-0.320	-0.088	-2.158	1
<i>Between Ex_Ch and Co_Ch:</i>					
W (C) - W (E)	-0.008	-0.122	0.106	-0.546	0
NW (C) - NW (E)	-0.028	-0.142	0.086	0.269	0
<i>Between Ex_Ad and Co_Ad:</i>					
W (C) - W (E)	-0.003	-0.117	0.110	0.418	0
NW (C) - NW (E)	-0.023	-0.137	0.090	-0.663	0
<i>Within Co_Ch:</i>					
NW - W	-0.062	-0.169	0.045	-0.106	0
<i>Within Co_Ad:</i>					
NW - W	0.099	-0.016	0.215	0.329	0
<i>Between Co_Ch and Co_Ad:</i>					
W (Ch) - W (Ad)	-0.047	-0.158	0.064	-0.766	0
NW (Ch) - NW (Ad)	-0.209	-0.320	-0.097	-1.352	1

### ***Reaction times – words-non-words; across groups, T1***

Next, we examined whether the mean transformed reaction times in the trials with the real words differed significantly from the ones with non-words.

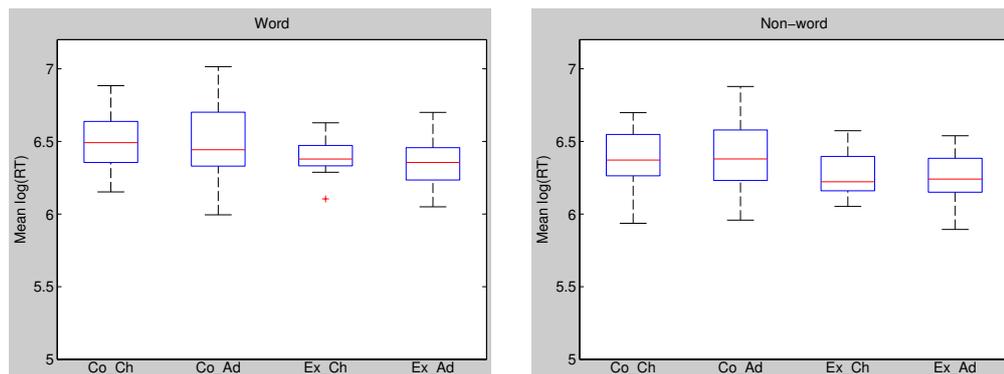


Figure 15: Boxplots to compare mean transformed reaction times, for words and non-words, for all four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1;  $n = 68$ ). Transformed chance level equals .79.

Figure 15 shows the comparison of the mean transformed reaction times (to correct trials) across the four groups for the comparison of words versus non-words. An ANOVA) was conducted, with Group (control vs. experimental), Age (adult vs.

child) and Word Type (word vs. non-word) as independent variables and Reaction Time as dependent variable, after Levene's test for equal variances was confirmed ( $F = 2.06, p = .053$ ). The ANOVA revealed a main effect of Group ( $F(1, 129) = 11.92, p = .0008$ ), and a main effect of Word Type ( $F(1, 129) = 54.56, p = .004$ ). There was no significant interaction of Group and Word Type and crucially, no main effect for Age ( $F(1, 129) = .009, p = .92$ ).

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007).

Table 15 summarises the results of the Tukey's test. The analyses showed no significant difference of mean transformed reaction times between words and non-words regardless whether the comparison was performed within or between any of the groups.

Table 15: Tukey's honestly significant difference criterion for mean transformed reaction times in the comparisons of words (W) and non-words (NW), across the four groups (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1).

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
<i>Within Ex_Ch:</i>					
NW - W	-0.134	-0.329	0.061	-0.871	0
<i>Within Ex_Ad:</i>					
NW - W	-0.085	-0.265	0.094	-0.567	0
<i>Between Ex_Ch and Ex_Ad:</i>					
W (Ch) - W (Ad)	0.039	-0.149	0.227	0.197	0
NW (Ch) - NW (Ad)	-0.010	-0.197	0.178	-0.001	0
<i>Between Ex_Ch and Co_Ch:</i>					
W (C) - W (E)	0.103	-0.082	0.288	0.639	0
NW (C) - NW (E)	0.116	-0.069	0.301	0.578	0
<i>Between Ex_Ad and Co_Ad:</i>					
W (C) - W (E)	0.126	-0.057	0.310	0.530	0
NW (C) - NW (E)	0.139	-0.045	0.322	0.652	0
<i>Within Co_Ch:</i>					
NW - W	-0.122	-0.295	0.051	-0.638	0
<i>Within Co_Ad:</i>					
NW - W	0.099	-0.016	0.215	-0.232	0
<i>Between Co_Ch and Co_Ad:</i>					
W (Ch) - W (Ad)	0.016	-0.164	0.196	0.103	0
NW (Ch) - NW (Ad)	-0.033	-0.213	0.148	-0.170	0

#### 4.4.5 Between groups – Comparison of non-word conditions (T1)

##### *Accuracy – non-words; between groups, T1*

In a next step, we compared the four groups in the four syllable conditions in terms of mean transformed hit rate.

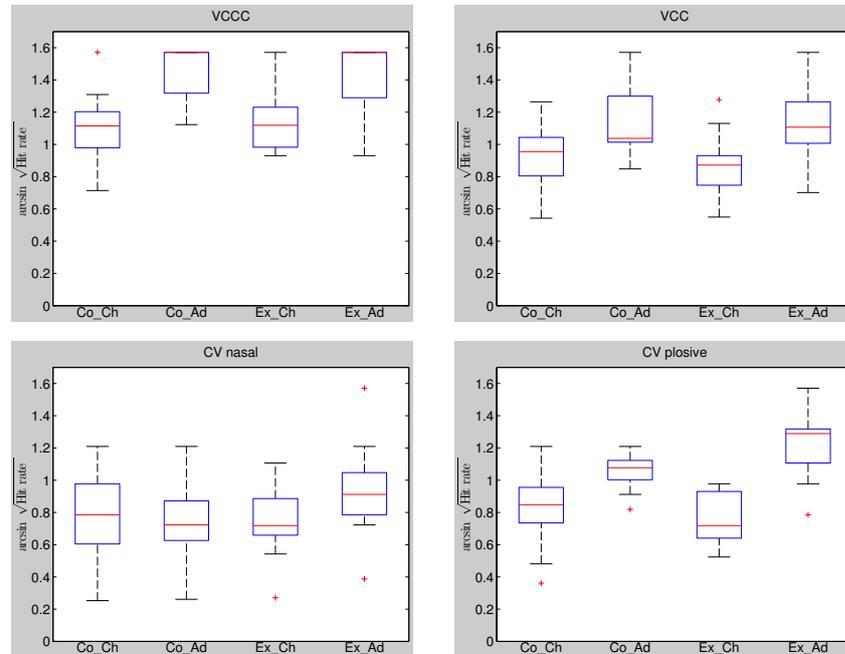


Figure 16: Boxplots to compare mean transformed hit rate per condition for each group (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1,  $n = 68$ ). Transformed chance level equals .79.

Figure 16 shows the comparison of the four groups for the four conditions. In both the Control and the Experimental group, adults had higher hit rates than children on all but one type (CV\_nasal). An omnibus ANOVA was conducted, with Group, Age and Syllable Condition as independent variables and Accuracy as dependent variable. A Levene's test for equal variances was not confirmed for the comparison of the four conditions ( $F = 1.75, p = .040$ ). The ANOVA revealed a main effect of Age ( $F(1, 463) = 165, p = .00001$ ), a main effect of Syllable Condition ( $F(3, 463) = 85.3, p = .0006$ ), and significant interactions of Group and Age ( $F(1, 463) = 11.9, p = .00001$ ), and of Age and Syllable Condition ( $F(3, 463) = 13.0, p = .00001$ ). There was no main effect for Group ( $F(1, 463) = 1.57, p = .21$ ).

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007).

Table 16: Tukey's honestly significant difference criterion for mean transformed hit rate between the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1).

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
<i>Within Ex_Ch:</i>					
VCC - VCCC	-0.252	-0.443	-0.061	-1.512	1
VCC - CV_nasal	0.061	0.095	0.217	0.579	0
VCC - CV_plosive	0.043	0.113	0.199	0.643	0
CV_plosive - CV_nasal	0.018	0.138	0.174	0.031	0
CV_nasal - VCCC	-0.313	-0.504	-0.122	-1.961	1
CV_plosive - VCCC	-0.295	-0.486	-0.104	-2.361	1
<i>Within Ex_Ad:</i>					
VCC - VCCC	-0.337	-0.513	-0.160	-1.608	1
VCC - CV_nasal	0.242	0.098	0.386	0.920	1
VCC - CV_plosive	-0.066	0.210	0.078	-0.600	0
CV_plosive - CV_nasal	0.308	0.164	0.452	1.389	1
CV_nasal - VCCC	-0.579	-0.755	-0.402	-2.164	1
CV_plosive - VCCC	-0.271	-0.447	-0.094	-0.902	1
<i>Between Ex_Ch and Ex_Ad:</i>					
VCCC (Ch) - VCCC (Ad)	-0.400	-0.583	-0.216	-1.344	1
VCC (Ch) - VCC (Ad)	-0.315	-0.454	-0.177	-1.430	1
CV_plosive (Ch) - CV_plosive (Ad)	-0.424	-0.562	-0.285	-2.650	1
CV_nasal (Ch) - CV_nasal (Ad)	-0.134	-0.272	0.005	-0.776	0
<i>Between Ex_Ch and Co_Ch:</i>					
VCCC (Co) - VCCC (Ex)	0.059	-0.124	0.241	-0.306	0
VCC (Co) - VCC (Ex)	0.101	-0.036	0.239	0.275	0
CV_plosive (Co) - CV_plosive (Ex)	0.011	-0.127	0.148	0.431	0
CV_nasal (Co) - CV_nasal (Ex)	-0.010	-0.147	0.127	0.142	0
<i>Between Ex_Ad and Co_Ad:</i>					
VCCC (Co) - VCCC (Ex)	-0.071	0.253	0.111	0.254	0
VCC (Co) - VCC (Ex)	-0.028	-0.164	0.108	0.097	0
CV_plosive (Co) - CV_plosive (Ex)	-0.119	-0.255	0.017	-1.087	0
CV_nasal (Co) - CV_nasal (Ex)	-0.140	-0.276	-0.003	-0.763	1
<i>Within Co_Ch:</i>					
VCC - VCCC	-0.209	-0.379	-0.039	-0.917	1
VCC - CV_nasal	0.172	0.033	0.311	0.588	1
VCC - CV_plosive	0.134	-0.005	0.272	0.450	0
CV_plosive - CV_nasal	0.039	-0.100	0.177	-0.209	0
CV_nasal - VCCC	-0.381	-0.551	-0.211	-1.320	1
CV_plosive - VCCC	-0.343	-0.513	-0.173	-1.390	1
<i>Within Co_Ad:</i>					
VCC - VCCC	-0.294	-0.477	-0.110	-1.717	1
VCC - CV_nasal	0.354	0.204	0.503	1.830	1
VCC - CV_plosive	0.025	-0.125	0.175	0.507	0
CV_plosive - CV_nasal	0.329	0.179	0.479	1.794	1
CV_nasal - VCCC	-0.647	-0.831	-0.464	-3.526	1
CV_plosive - VCCC	-0.319	-0.502	-0.135	-3.422	1
<i>Between Co_Ch and Co_Ad:</i>					
VCCC (Ch) - VCCC (Ad)	-0.270	-0.451	-0.089	-2.139	1
VCC (Ch) - VCC (Ad)	-0.186	-0.320	-0.051	-1.142	1
CV_plosive (Ch) - CV_plosive (Ad)	-0.294	-0.429	-0.159	-1.491	1
CV_nasal (Ch) - CV_nasal (Ad)	-0.004	-0.139	0.131	0.150	0

Tukey's HSD criterion showed different pictures for syllable classification in terms of mean transformed hit rate in the different groups (cf. Table 16). Experimental children classified VCCC significantly better than VCC, CV\_nasal and CV\_plosive,

but VCC was not classified significantly different from CV\_nasal or from CV\_plosive, and CV\_nasal was not significantly different from CV\_plosive. For experimental adults, on the other hand, the same pattern arose as we have seen in Study 1. Each syllable condition was significantly different from every other syllable condition in terms of mean hit rate, except for CV\_plosive from VCC. The comparison of the four syllable conditions between the Experimental Children and Adults showed that the two groups had significantly different hit rates on all syllable conditions but CV\_nasal. The comparison of the four syllable conditions between the Experimental Children and Control Children showed no significant differences in any of the syllable conditions. The comparison of the four syllable conditions between the Experimental Adults and the Control Adults, however, showed a significant difference in the CV\_nasal condition. The comparisons within the Control groups mirrored the Experimental groups completely, safe for the comparison of VCC and CV\_nasal that differed in the Control Children (but not in the Experimental Children).

***Reaction times– non-words; between groups, T1***

In a next step, we compared the four groups in the four syllable conditions in terms of mean transformed reaction times.

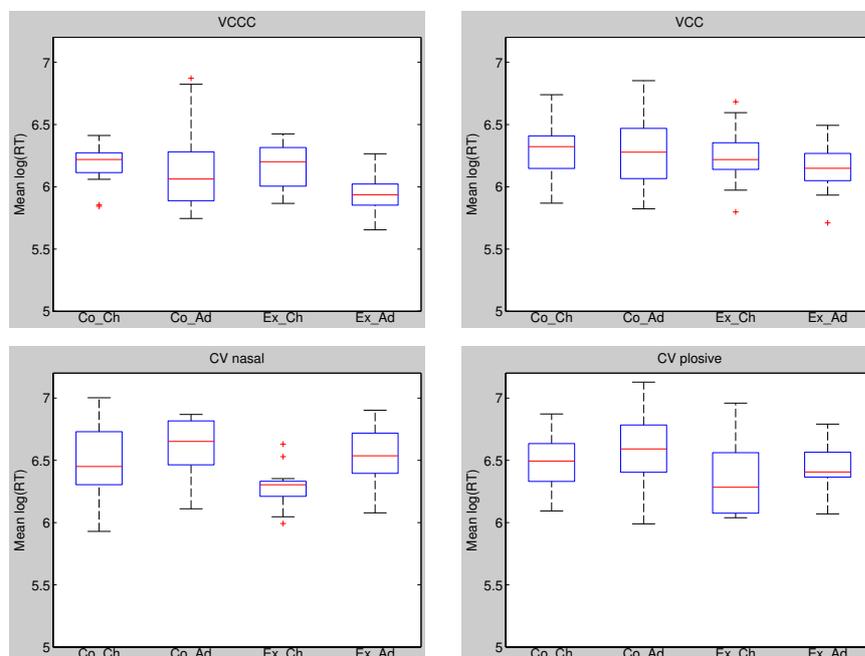


Figure 17: Boxplots to compare mean transformed reaction times, per condition for each group (Co\_Ch = Control Children, Co\_Ad = Control Adults, Ex\_Ch = Experimental Children time 1, Ex\_Ad = Experimental Adults time 1,  $n = 68$ ).

Figure 17 shows the comparison of the mean transformed reaction times (to correct trials) across the four groups in the four conditions. An omnibus ANOVA was conducted, with Group, Age and Syllable Condition as independent variables and Reaction Time as dependent variable. A Levene's test for equal variances was not confirmed for the comparison of four conditions ( $F = 3.28, p = .00003$ ). The ANOVA revealed a main effect of Group ( $F(1, 268) = 29.1, p = .00001$ ), a main effect of Syllable Condition ( $F(3, 268) = 54.6, p = .00001$ ), and a significant interaction for Age and Syllable Condition ( $F(3, 268) = 7.23, p = .0001$ ). There was no significant interaction of Group and Syllable Condition and crucially, no main effect for Age.

Post-hoc multiple comparison tests were carried out to explore the main effect with Tukey's honestly significant difference (HSD) criterion (Linton & Harder, 2007).

Tukey's HSD criterion showed different pictures for syllable classification in terms of mean transformed reaction times in the different groups (cf. Table 17). Experimental children only classified VCCC significantly faster than CV\_plosive; all other contrasts did not reach significance. Control children differed from the experimental children in three more syllable contrasts that reached significance (VCC-CV\_nasal, VCC-CV\_plosive, and CV\_nasal-VCCC). Experimental adults, on the other hand, classified VCCC and VCC each significantly faster than CV\_nasal and CV\_plosive, but mean reaction times did not differ significantly between conditions VCCC and VCC nor between conditions CV\_plosive and CV\_nasal. Control adults performed like the experimental adults for all syllable comparisons. Experimental children and experimental adults showed no significant group difference for mean reaction times in any condition, and neither did experimental children and control children. Control children and control adults differed in mean reaction times in condition CV\_nasal. Experimental adults and the control adults, however, showed a significant difference in mean reaction times in condition CV\_nasal.

Table 17: Tukey's honestly significant difference criterion for mean transformed reaction times between the four conditions (VCCC, VCC, CV\_plosive and CV\_nasal) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1).

	Difference	Conf_left	Conf_right	Cohen's d	Reject H0
<i>Within Ex_Ch:</i>					
VCC - VCCC	0.112	-0.117	0.341	0.388	0
VCC - CV_nasal	-0.101	-0.288	0.086	-0.180	0
VCC - CV_plosive	-0.130	-0.318	0.057	-0.402	0
CV_plosive - CV_nasal	0.029	-0.158	0.216	0.294	0
CV_nasal - VCCC	0.214	0.016	0.443	0.689	0
CV_plosive - VCCC	0.243	0.014	0.472	0.710	1
<i>Within Ex_Ad:</i>					
VCC - VCCC	0.201	-0.011	0.412	1.248	0
VCC - CV_nasal	-0.306	-0.479	-0.134	1.640	1
VCC - CV_plosive	-0.258	-0.430	0.085	1.630	1
CV_plosive - CV_nasal	-0.049	-0.221	0.124	0.390	0
CV_nasal - VCCC	0.507	0.296	0.718	2.568	1
CV_plosive - VCCC	0.459	0.247	0.670	3.105	1
<i>Between Ex_Ch and Ex_Ad:</i>					
VCCC (Ch) - VCCC (Ad)	0.138	-0.082	0.358	1.275	0
VCC (Ch) - VCC (Ad)	0.050	-0.116	0.216	0.434	0
CV_plosive (Ch) - CV_plosive (Ad)	-0.078	-0.244	0.088	-0.360	0
CV_nasal (Ch) - CV_nasal (Ad)	-0.155	-0.321	0.010	-1.082	0
<i>Between Ex_Ch and Co_Ch:</i>					
VCCC (Co) - VCCC (Ex)	0.114	-0.105	0.333	0.167	0
VCC (Co) - VCC (Ex)	0.077	-0.087	0.242	0.174	0
CV_plosive (Co) - CV_plosive (Ex)	0.132	-0.033	0.296	0.514	0
CV_nasal (Co) - CV_nasal (Ex)	0.147	-0.017	0.311	0.883	0
<i>Between Ex_Ad and Co_Ad:</i>					
VCCC (Co) - VCCC (Ex)	0.134	-0.085	0.352	0.788	0
VCC (Co) - VCC (Ex)	0.097	-0.067	0.260	0.556	0
CV_plosive (Co) - CV_plosive (Ex)	0.151	-0.012	0.314	0.592	0
CV_nasal (Co) - CV_nasal (Ex)	0.167	0.003	0.330	0.398	1
<i>Within Co_Ch:</i>					
VCC - VCCC	0.075	-0.128	0.279	0.445	0
VCC - CV_nasal	-0.171	-0.337	-0.005	-0.805	1
VCC - CV_plosive	-0.185	-0.351	-0.019	-0.845	1
CV_plosive - CV_nasal	0.014	-0.152	0.180	-0.057	0
CV_nasal - VCCC	0.246	0.043	0.450	1.205	1
CV_plosive - VCCC	0.260	0.057	0.464	1.383	1
<i>Within Co_Ad:</i>					
VCC - VCCC	0.164	-0.056	0.384	0.443	0
VCC - CV_nasal	-0.376	-0.556	-0.197	-1.247	1
VCC - CV_plosive	-0.312	-0.492	-0.133	-0.954	1
CV_plosive - CV_nasal	-0.064	-0.244	0.116	-0.101	0
CV_nasal - VCCC	0.540	0.320	0.760	1.676	1
CV_plosive - VCCC	0.476	0.256	0.696	1.285	1
<i>Between Co_Ch and Co_Ad:</i>					
VCCC (Ch) - VCCC (Ad)	0.119	-0.098	0.335	0.147	0
VCC (Ch) - VCC (Ad)	0.030	-0.131	0.192	-0.014	0
CV_plosive (Ch) - CV_plosive (Ad)	-0.097	-0.259	0.064	-0.334	0
CV_nasal (Ch) - CV_nasal (Ad)	-0.175	-0.336	-0.013	-0.397	1

#### 4.4.6 T1 versus T2 – adults’ versus children’s performance on non-words

##### *Accuracy – non-words; T1 versus T2*

Next, we compared the performance of the Experimental group in terms of mean transformed hit rates at time 1 (T1) and time 2 (T2).

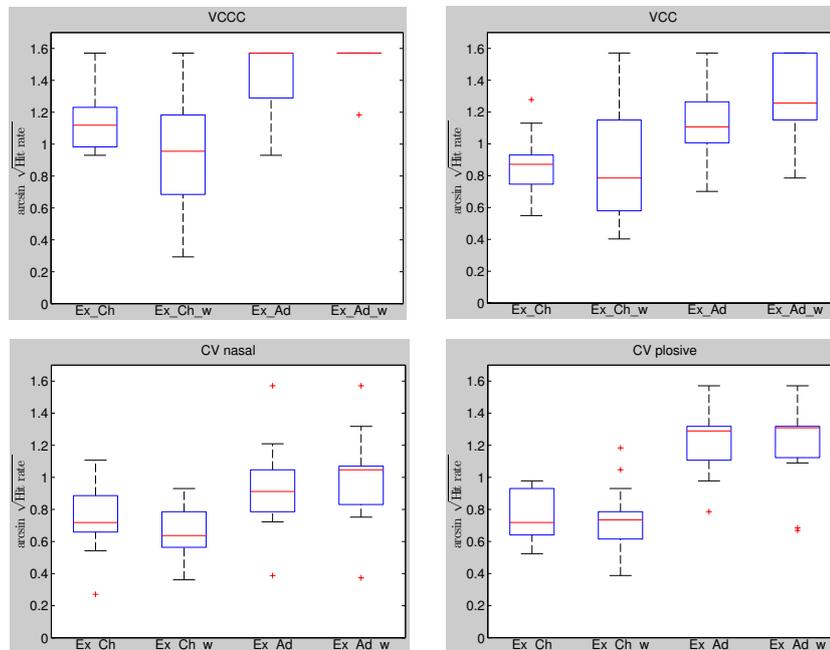


Figure 18: Boxplots to compare the mean transformed hit rate per condition, comparing the two experimental groups at T1 and T2 (Ex\_Ch = Experimental Children T1, Ex\_Ch\_w = Experimental Children T2, Ex\_Ad = Experimental Adults T1, Ex\_Ad\_w = Experimental Adults T2). Ex\_Ch/Ex\_Ch\_w:  $n = 14$ ; Ex\_Ad/Ex\_Ad\_W:  $n = 18$ . Transformed chance level equals .79.

Figure 18 shows the comparison of the experimental group (Experimental Children, Experimental Adults) at time 1 and time 2. Separate repeated-measures ANOVAs<sup>33</sup> per Syllable Condition, with Time and Age as the independent and Accuracy as the dependent variables, revealed a main effect of Age ( $F(1, 30) = 37.5, p = .00001$ ) and a significant interaction of Age and Time ( $F(1, 30) = 8.47, p = .007$ ) for condition VCCC; a main effect of Age ( $F(1, 30) = 22.6, p = .00001$ ) for condition VCC; a main effect of Age ( $F(1, 30) = 12.7, p = .001$ ) for condition CV\_nasal; and a main effect of Age ( $F(1, 30) = 60.6, p = .00001$ ) for condition CV\_plosive. There was no effect of Time only in any of the syllable conditions, which indicates that performance did not change over time.

<sup>33</sup> Levene’s test was confirmed for types VCC ( $F = 3.78, p = 0.015$ ), CV\_nasal ( $F = 0.49, p = 0.69$ ), and CV\_plosive ( $F = 0.55, p = 0.65$ ), but not for type VCCC ( $F = 9.15, p = 0.00005$ ).

### Reaction times – non-words; T1 versus T2

Next, we compared the performance of the Experimental group in terms of mean transformed reaction times at time 1 (T1) and time 2 (T2).

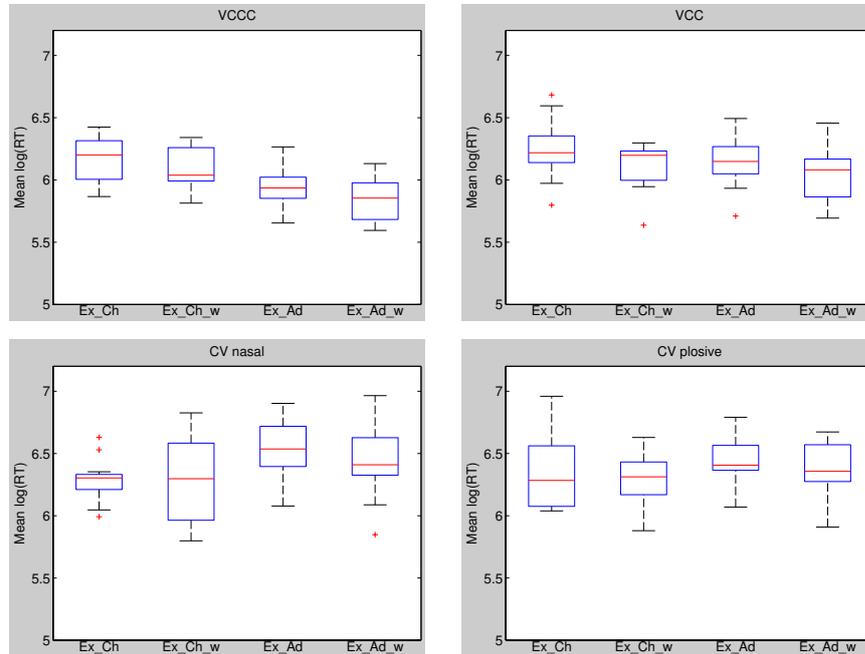


Figure 19: Boxplots to compare the mean transformed reaction times per condition, comparing the two experimental groups at T1 and T2 (Ex\_Ch = Experimental Children T1, Ex\_Ch\_w = Experimental Children T2, Ex\_Ad = Experimental Adults T1, Ex\_Ad\_w = Experimental Adults T2). Ex\_Ch/Ex\_Ch\_w:  $n = 14$ ; Ex\_Ad/Ex\_Ad\_W:  $n = 18$ .

Figure 19 shows the comparison of the Experimental group (Experimental Children, Experimental Adults) at time 1 and time 2. Repeated-measures ANOVAs<sup>34</sup> for each Syllable Condition, with Time and Age as the independent and Reaction Time as the dependent variables, revealed a main effect of Time ( $F(1, 30) = 6.69, p = .015$ ), a main effect of Age ( $F(1, 30) = 20.40, p = .0001$ ) (but no significant interaction of Time and Age ( $F(1, 30) = .02, p = .89$ )) for condition VCCC; a main effect of Time ( $F(1, 30) = 8.80, p = .006$ ) for condition VCC; a main effect of Age ( $F(1, 30) = 5.80, p = .022$ ) for condition CV\_nasal; and no main effect of Age ( $F(1, 30) = 1.88, p = .18$ ) for condition CV\_plosive.

<sup>34</sup> Levene's test was confirmed for all four types: VCCC ( $F = 0.53, p = 0.66$ ), VCC ( $F = 0.73, p = 0.43$ ), CV\_nasal ( $F = 1.74, p = 0.17$ ), CV\_plosive ( $F = 2.11, p = 0.11$ ).

#### 4.4.7 Within subjects comparisons (T1 versus T2)

##### *Accuracy – non-words; within subject, T1 versus T2*

Figure 20 shows the within-subject comparisons for the mean transformed hit rate in each syllable condition at time 1 (T1) and time 2 (T2) for each Adult in the Experimental group, and Figure 21 shows the same for Experimental Children.

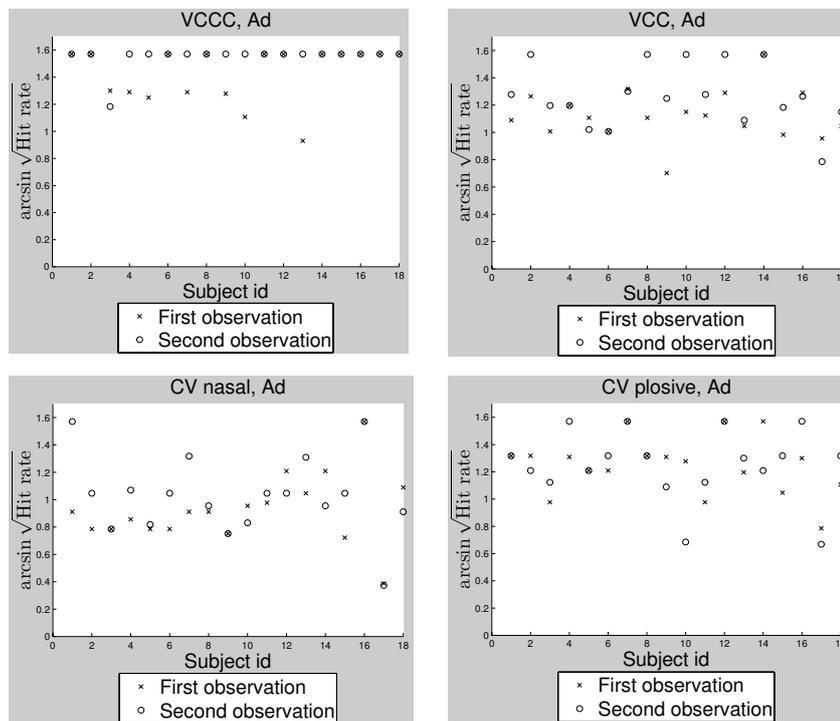


Figure 20: Individual subject comparisons per condition of mean transformed hit rate T1 versus T2 (Ad = Experimental Adults,  $n = 18$ ). Transformed chance level equals .79.

Figure 20 illustrates that most adults performed more accurately at T2 than at T1. In condition VCCC, only one participant did not improve, the rest performed at ceiling at T2 (12 already performed at ceiling at T1). Except for two participants, everyone performed better at T2 in condition VCC. 13 out of 18 participants performed better at T2 in condition CV\_plosive and 14 out of 18 in condition CV\_nasal.

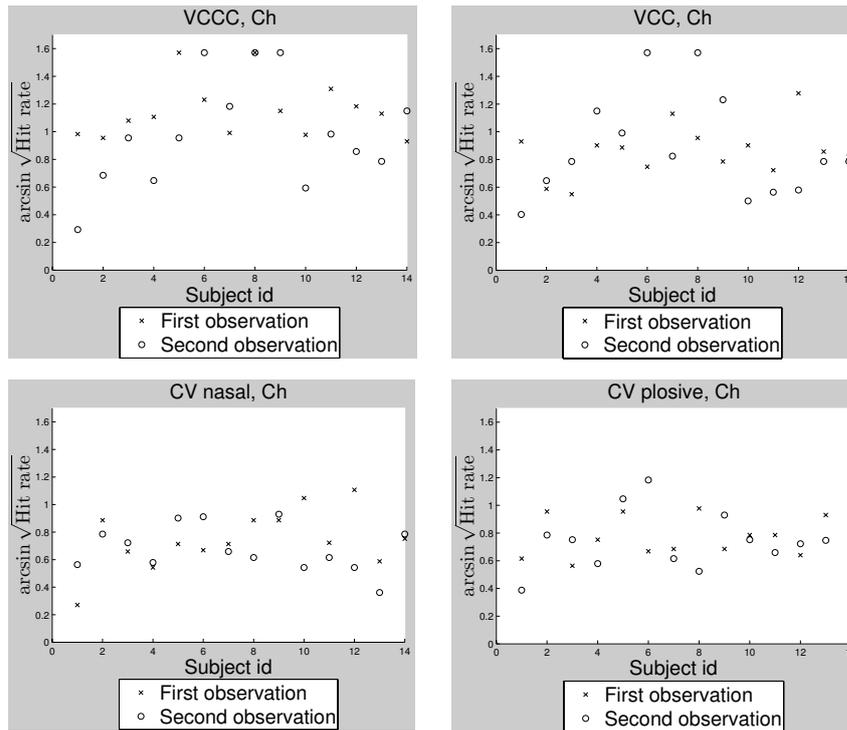


Figure 21: Individual subject comparisons per condition of mean transformed hit rate T1 versus T2 (Ch = Experimental Children,  $n = 14$ ). Transformed chance level equals .79.

Figure 21 shows that children’s accuracy differed more randomly between T1 and T2 than for adults. In condition VCCC, only 4 out of 14 performed better at T2 than at T1. Every other child performed better at T2 in condition VCC, and 6 out of 14 children performed better at T2 in both conditions CV\_plosive and CV\_nasal.

***Reaction times – non-words; within subject, T1 versus T2***

Figure 22 shows the within-subject comparisons for the mean transformed reaction times in each syllable condition at time 1 (T1) and time 2 (T2) for each adult in the Experimental group, and Figure 23 shows the same for the Experimental Children.

Figure 22 illustrates that adults’ mean transformed reaction times were generally shorter at time 2. Overall, however, the inter-individual differences were small. In condition VCCC, 13 out of 18 adults performed faster at T2. 14 out of 18 performed faster at T2 in both condition VCC and CV\_nasal. In condition CV\_plosive, 11 out of 18 performed faster at T2. Reaction times were fastest in condition VCCC and slowest in CV\_nasal and CV\_plosive.

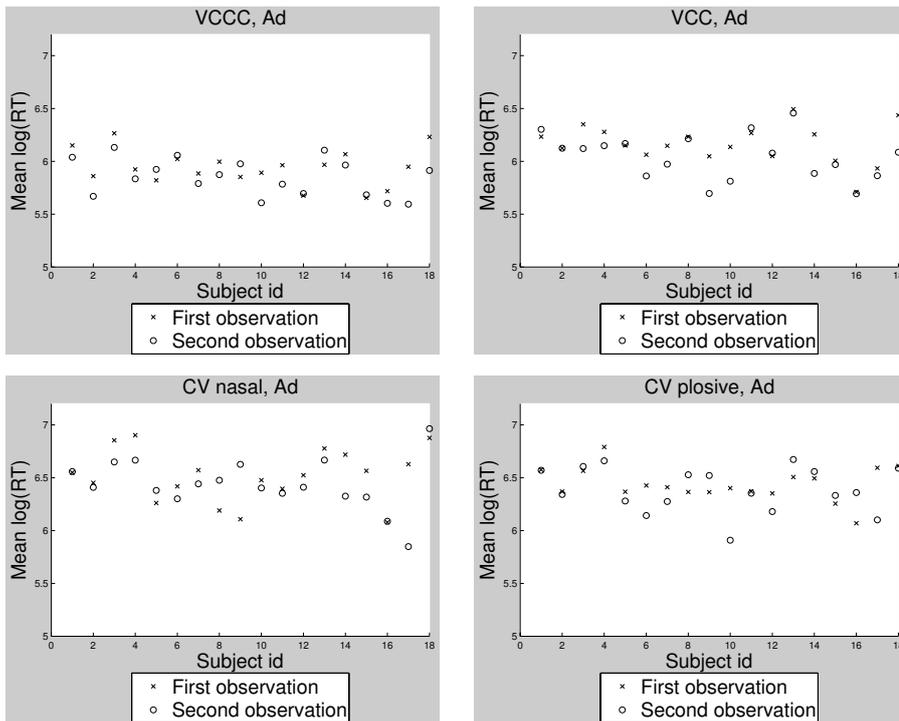


Figure 22: Individual subject comparisons per condition of mean transformed reaction times T1 versus T2 (Ad = Experimental Adults, n = 18).

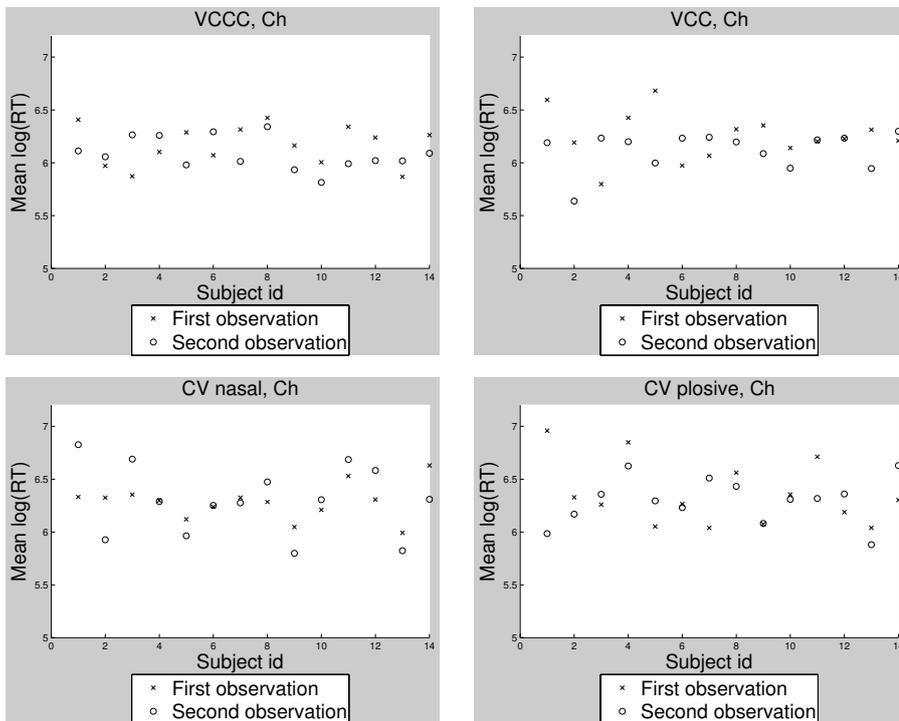


Figure 23: Individual subject comparisons per condition of mean transformed reaction times T1 versus T2 (Ch = Experimental Children, n = 14).

Figure 23 shows that children’s overall mean transformed reaction times varied more strongly inter-individually than in adults. In condition VCCC, 11 out of 14 children performed faster at T2. 10 out of 14 children performed faster at T2 in condition VCC. In condition CV\_plosive, 9 out of 14 performed faster at T2, and 8 out of 14 in CV\_nasal. It is noticeable that Children at T2 showed much more variation in mean reaction times in condition CV\_nasal than at time 1.

#### 4.4.8 Correlations with cognitive variables

In a next step, we performed analyses to investigate whether performance on the LDT correlated with measures on the cognitive variables. For the correlation of working memory across the lifespan, we expected developments according to the literature (cf. Baltes, 1987). For the correlation of working memory and the number of foreign languages with the four syllable conditions, we expected the strongest correlations with syllable-condition VCCC, and VCC, and less strong with CV\_plosive and least strong with CV\_nasal.

##### *Experimental group*

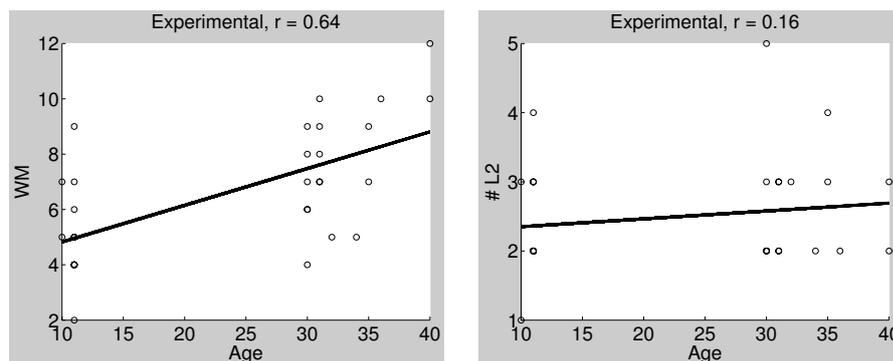


Figure 24: Linear correlation between age and cognitive variables; (left) working memory (WM); (right) number of foreign languages without German (#L2),  $n = 32$ .

Figure 24 shows the correlation of the two cognitive variables we tested in Study 2 for the experimental groups (at time 1) and age. Contrary to results obtained in Study 1 (cf. Section 3.4.6), there was a significant correlation of working memory (WM) and age ( $r(32) = .64$ ,  $p < .001$ ). In contrast, there was no significant correlation with the number of foreign languages (#L2s) and age ( $r(32) = .16$ ,  $p > .05$ ).

Next, we examined whether responses to individual syllable conditions in terms of mean transformed hit rates correlated with performance on the cognitive variables.

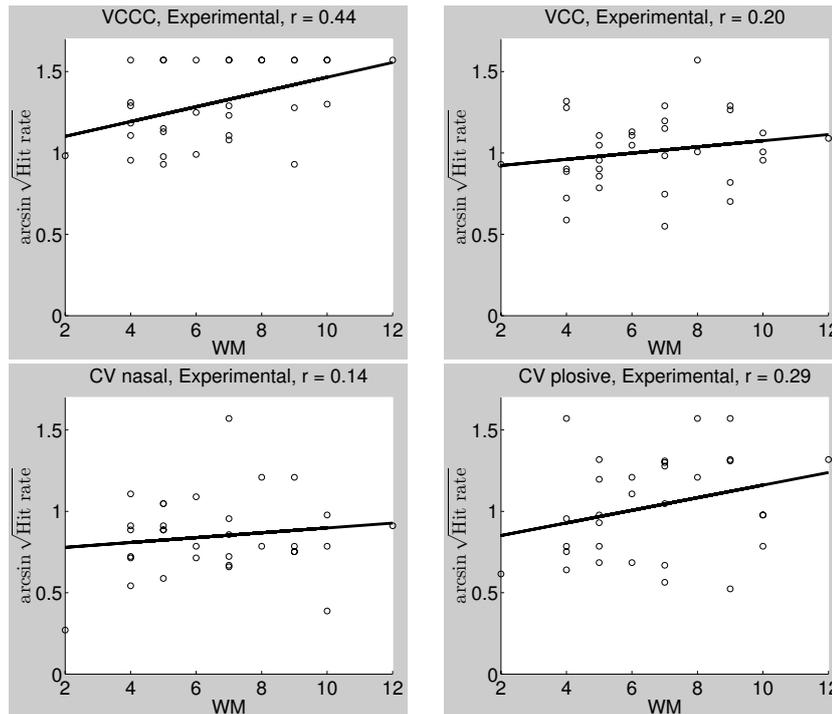


Figure 25: Linear correlation between working memory and transformed hit rate for each experimental condition,  $n = 32$ . Transformed chance level equals .79.

Figure 25 shows the correlation between working memory and transformed hit rates for each syllable condition. WM correlated significantly with hit rate in condition VCCC ( $r(32) = .44, p < .05$ ). However, the correlation did not reach significance in any other syllable condition: VCC ( $r(32) = .20, p > .05$ ), CV\_plosive ( $r(32) = .29, p > .05$ ), and CV\_nasal ( $r(32) = .14, p > .05$ ).

Figure 26 shows the correlations between the number of foreign languages (German not included) and transformed hit rates for each syllable condition. None of the correlations reached significance (VCCC ( $r(32) = .21, p > .05$ ), VCC ( $r(32) = .24, p > .05$ ), CV\_nasal ( $r(32) = .25, p > .05$ ), CV\_plosive ( $r(32) = .30, p > .05$ )).

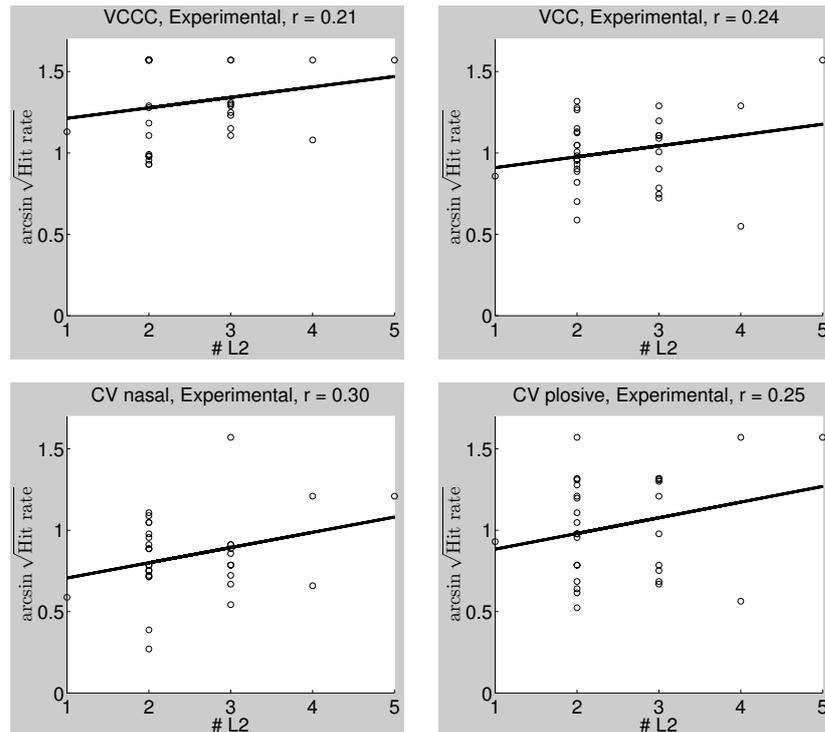


Figure 26: Linear correlation between number of foreign languages and transformed hit rate for each experimental condition,  $n = 32$ . Transformed chance level equals .79.

**Control group**

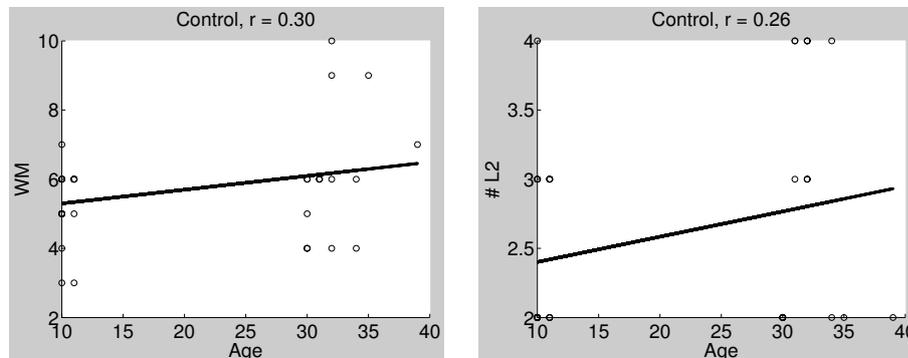


Figure 27: Linear correlation between age and cognitive variables; (left) working memory (WM); (right) number of foreign languages without German (#L2),  $n = 36$ .

Figure 27 shows the correlation of the two cognitive variables for the control groups and age. In contrast to the experimental group, there was no significant correlation, neither with working memory ( $r(36) = .30, p > .05$ ), nor with the number of foreign languages and age ( $r(36) = .26, p > .05$ ).

Next, we examined whether responses to individual syllable conditions in terms of mean transformed hit rates correlated with the performance of the cognitive variables.

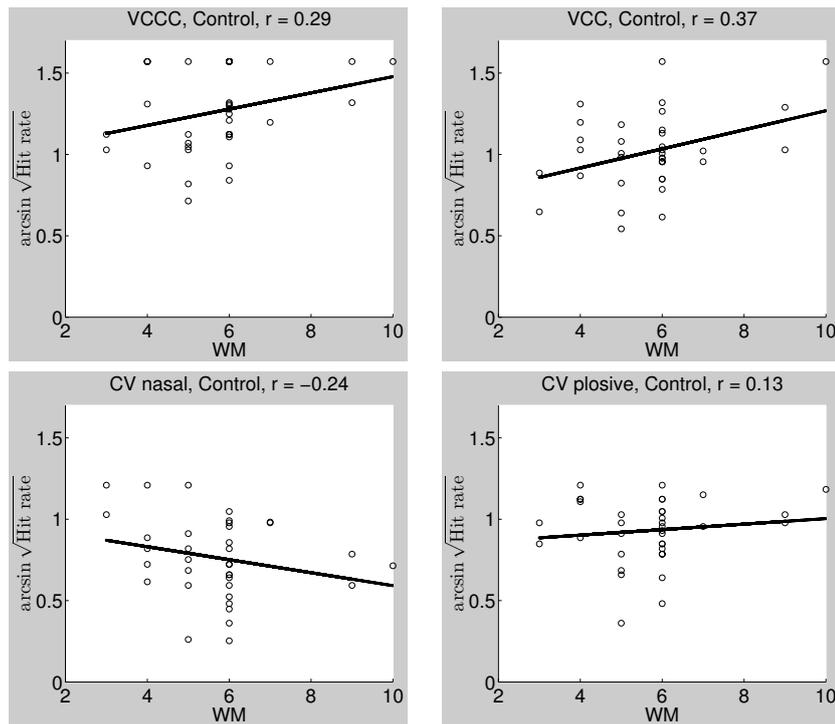


Figure 28: Linear correlation for working memory and transformed hit rate for each experimental condition,  $n = 36$ . Transformed chance level equals .79.

Figure 28 shows the correlation between working memory and transformed hit rates (in arcsine-square-root) for each syllable condition. WM correlated significantly with hit rate in condition VCC ( $r(36) = .37, p < .05$ ). However, the correlation did not reach significance in any other syllable condition (VCCC ( $r(36) = .29, p > .05$ ), CV\_plosive ( $r(36) = .13, p > .05$ ), CV\_nasal ( $r(36) = -.24, p > .05$ )).

Figure 29 shows the correlation between the number of foreign languages (German not included) and transformed hit rates for each syllable condition. The number of L2s correlated significantly with hit rate in condition VCC ( $r(36) = .52, p < .01$ ). The correlation, however, did not reach significance in any other syllable condition (VCCC ( $r(36) = .27, p > .05$ ), CV\_plosive ( $r(36) = .10, p > .05$ ), CV\_nasal ( $r(36) = -.10, p > .05$ )).

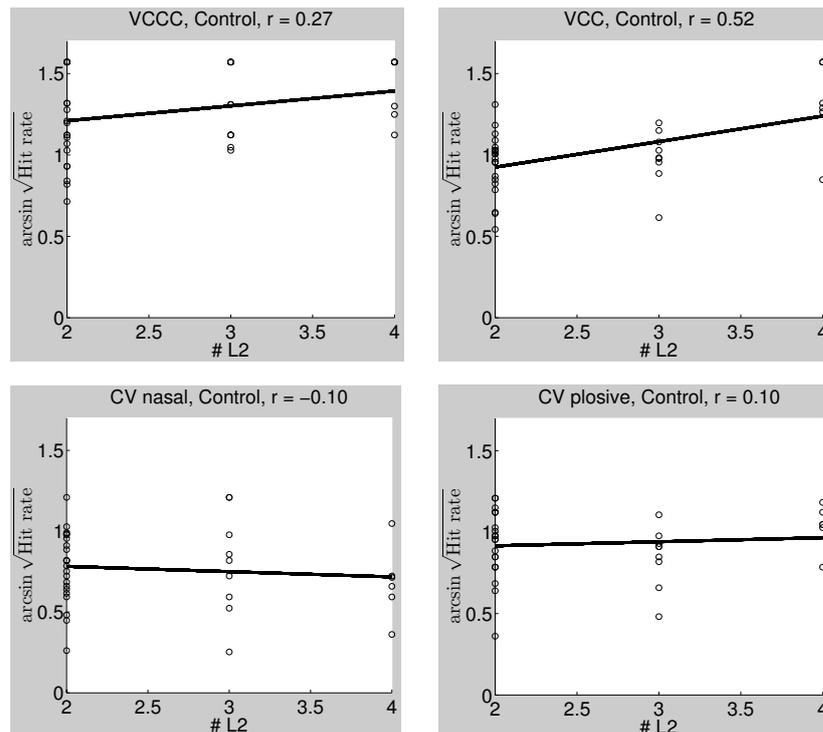


Figure 29: Linear correlation for number of foreign languages and transformed hit rates for each experimental condition,  $n = 36$ . Transformed chance level equals .79.

#### 4.4.9 Summary

The results revealed that the Experimental groups only partially outperformed the Control groups. There was no significant difference (on neither accuracy nor reaction times) between the Child Control and Child Experimental groups. However, adults in the Experimental group differed significantly (on both accuracy and reaction times) from adults in the Control group for condition CV\_nasal. For the distinction of words and non-words, adults and children performed equally well on the words, but adults outperformed children on the non-words in terms of accuracy. There were no significant differences between adults and children in terms of reaction time. Neither Control and Experimental children nor Control and Experimental adults differed in either words or non-words (on neither accuracy nor reaction time).

In terms of accuracy, Experimental adults performed similarly to Study 1 (cf. Section 3.4.7). Adults classified some syllable conditions better than others (VCCC > VCC = CV\_plosive > CV\_nasal; where ‘>’ means higher accuracy and ‘=’ no significant difference). Similar results were also found for Experimental children (VCCC > VCC = CV\_plosive = CV\_nasal). In terms of reaction time, Experimental adults only partly performed according to the pattern that arose in Study 1 (VCCC =

VCC > CV\_plosive = CV\_nasal; where '>' means higher reaction time and '=' no significant difference). Slightly different relations were found for Experimental children (VCCC > CV\_plosive; VCCC = VCC = CV\_plosive = CV\_nasal). We observed no significant effects between Control and Experimental Children groups. However, adults in the Experimental group were both better and faster at classifying one syllable condition (CV\_nasal) indicating that a learning process indeed took place.

Accuracy and reaction times remained stable after one week indicating the perseverance of learning. In terms of accuracy, adults outperformed children in all four conditions, at time 1 and time 2. In terms of reaction time, adults and children were able to benefit from the consolidation and exhibited faster reaction times on two syllable conditions (VCCC and VCC) at time 2.

Overall, Study 2 replicated the findings of study 1 in terms of mean accuracy and reaction time (VCCC > VCC > CV\_plosive > CV\_nasal). However, there were differences in correlations related to the cognitive variables. Study 2 examined a limited scope of age bands. Since the older age groups were not represented, there were no indications of a plateau in measurements of working memory, which increased with age in Study 2. Analogously, the number of L2s increased with age in Study 1 while it did not correlate significantly in Study 2. Results of Study 1 suggest that the elderly tend to know more languages, but they were not examined in Study 2.

## 4.5 Discussion

Study 2 asked three questions. The first question was whether there was evidence of real learning from input or whether general inferencing is enough to solve the Lexical Decision Task. Second, we tested whether children differed from adults in this regard. Finally, we also examined whether the capacity to distinguish words from non-words and the ability to generalize knowledge from the input to new stimuli changed after a consolidation period of one week or not. As in Study 1, we analyzed whether cognitive variables such as working memory and the number of foreign languages a person ‘knows’ affected the ability to distinguish ‘legal from illegal’ phonotactics in an unknown language. Unlike in Study 1, working memory correlated positively with age and some syllable conditions in the experimental group, but the number of L2s did not. Concerning the first question, we found evidence of real learning in the adult group that received input, given that they performed significantly better than the control group in syllable condition CV\_nasal. This finding differs from Study 1 where participants as a group performed at chance for the syllable condition CV\_nasal. Adults significantly differed from children what concerned learning from input. Children in both the experimental and the control group seem to have used general inferencing strategies, since there was no significant difference for any syllable condition between the children who had received input prior to the LDT and the ones that did not. The performance of both adults and children in the experimental group remained constant after the consolidation period of one week.

The results of Study 2 revealed differences between adults and children that could not be revealed in Study 1 due to the design of the study. In Study 2, age was represented in two groups instead of as a continuous variable. This allowed us to make group and age comparisons. Furthermore, the control group enabled us to ascribe differences in performance more specifically to the influence of the exposure or to the age variable. The possibility to re-test the experimental group allowed us to examine whether the learning effects were only transitory or became embedded in memory. Moreover, we could compare adults and children in this quasi-longitudinal regard.

#### **4.5.1 Effects of input or of previous intake?**

The finding that all our participants (including controls and participants in Study 1) correctly rejected the three- and two-consonant cluster syllables, is consistent with Roberts et al.'s findings (2010). Since we split up the CVC syllables into CV\_nasal and CV\_plosive syllables, we could additionally show that CV\_plosive syllables could also be rejected significantly above chance without any exposure to the WR. In a future study, we would like to investigate whether twice as much input (e.g. 14 minutes – like examined in Roberts et al. (2010)) would suffice for children to also perform above chance in the CV\_nasal condition, or how much input it would require to do so. At any rate, results from Study 2 confirm the results from Study 1, showing that there is a trend of better performance on the Lexical Decision Task with increasing age. Although we have to be cautious in interpreting null results, we conclude that the control groups and the experimental child group (and participants in Study 1) were able to discriminate words from non-words to some extent, but without relying on implicit learning mechanisms. Input only served experimental adults to significantly enhance the discrimination ability for the most difficult CV\_nasal syllables (cf. also Hayes-Harb, 2007).

Our finding that only adults seem to have learned something from input is in line with findings by, for example, Muñoz (2006) and Dimroth (2012), who found that adults were faster learners at the beginning of foreign language learning. Adults have been found to initially process language more efficiently than children, whereas children have been found to perform better at the rate and degree of retention or even generally better than adults (Carey & Bartlett, 1978). Keeping in mind that our exposure/learning time only lasted seven minutes, adults in our study clearly outperformed children.

Our results are further consistent with findings showing that more experience with the first language helps to make distinctions of L1 and L2 language sounds (e.g. Best et al., 2001). The finding that adults in the control group performed better than children in the VCCC, VCC and even in the CV\_plosive condition further supports the explanation of more language experience to be beneficial for the distinction of new stimuli of an unknown language, because VCCC and VCC are supposed to be most similar to L1 phonotactics. This counters studies that have found L1 phonology to be a constraint for the perception of L2 phonetic contrasts (e.g. Cutler, 2001; Cutler & Otake, 1994; Flege & Wang, 1989). Nittrouer (2004) found that adults differed from children in the perception of final stops. Supposedly, children favour

different signal properties in perceptual decisions than adults. This suggests that, additionally to different rates of learning, children might also use different routes.

#### **4.5.2 Consolidation effects?**

Comparing the performance at time one to time two, we found no significant difference in any syllable-condition for time in terms of accuracy. Therefore, we can conclude that the ‘knowledge’ that was gained through the brief exposure remained stable after the consolidation period of one week. This speaks against Markson and Bloom’s (1997) finding that adults were better at time 1 but that they lost their advantage after one week of retention period. It is important to note that Markson and Bloom’s (1997) learning task was an explicit one, whereas we tested implicit learning. Our result is in line with other studies on consolidation (cf. e.g. Bisson et al., 2013; Carey & Bartlett, 1978; Tamminen & Gaskell, 2006; Tamminen et al., 2010). Tamminen and Gaskell (2006) interpreted their findings of a lexicalization delay as evidence that information had been transferred from short- to long-term storage (cf. p. 827). Rapid incidental vocabulary learning was enhanced after a consolidation period of one week in the study of Bisson et al. (2013), both in terms of accuracy and in terms of reaction time. We only found differences in terms of reaction time between time one and time two, but no increased accuracy like in Bisson et al. (2013).

In our study, we could not distinguish the effect of sleep from the effect of a retention time (with or without sleep) like Tamminen et al. (2010) did. The authors concluded that sleep, not retention time, is crucial for the consolidation of explicit knowledge. Fischer, Drosopoulos, Tsen and Born (2006), for example, provided evidence for the enhancement of implicit memory through sleep consolidation. Again, implicit memory in adults did not seem to be enhanced at time two, but, importantly, it also did not decrease. For children, we cannot speak of implicit learning because their performance was not significantly different from the no-input (control) group. Fischer et al. (2007) found that implicit knowledge in 7-11-year old children decreased after sleep retention. Even though they used a different implicit learning task (serial reaction time), their finding of sleep-dependent deterioration in children reinforces our null result of implicit learning in children, because our children performed equally well at time two (cf. also Diekelmann et al., 2009; Wilhelm et al., 2008). In the case of adults, we are potentially able to speak of resistance to deterioration at time two, since subjects did not receive any more ‘correct’ input at time two and the Lexical Decision Task at time one and two could

actually threaten the learning effect as it confronted subjects with 50% ‘bad’/ ‘incorrect’ input.

### **4.5.3 Cognitive or multilingual advantage?**

#### ***Working memory***

In Study 2, the only cognitive variable measured was working memory, using the same task that was used in Study 1. While we detected no significant correlation of WM with age in Study 1 (cf. Section 3.4.6), WM correlated significantly with age in the experimental group in Study 2 (cf. Section 4.4.8). Interestingly, WM did not correlate significantly with age in the control group. A possible reason for these differences could be the differences in group-size and/or the presence of ‘outliers’ (cf. e.g. Student, 1908; Wilcox, 2005). Since groups in Study 2 were substantially smaller (and the age-spans between 11-29 and 40-86 years were not covered) compared to the group in Study 1, small variations in WM scores could have led to big differences in correlation. Surprisingly, WM increased across the lifespan – a correlation that pointed in the opposite direction of findings of a linear decrease (cf. e.g. Borella et al., 2008; Jenkins et al., 1999; Park et al., 2002; Park & Payer, 2006). The experimental and the control group subsequently also differed in the correlation of WM with accuracy in the different syllable conditions. In the experimental group, WM correlated with condition VCCC, and in the control group with condition VCC, but both correlations were weak. Because there were no correlations of WM with the two critical conditions CV\_nasal and CV\_plosive, our results cannot concur with findings of WM as a significant predictor of learning (cf. e.g. Alloway, 2009).

#### ***Number of L2s***

Concerning the number of L2s, there was no significant correlation with age as in Study 1 (cf. Section 3.4.6), neither in the control nor in the experimental group (cf. Section 4.4.8). In the experimental group, the number of L2s did not correlate with accuracy in any of the syllable conditions. In the control group, the number of L2s correlated with accuracy in condition VCC only. It is difficult to find an explanation for the finding that the number of L2s correlated with accuracy in condition VCC but not in VCCC, especially taking into account that the correlations in Study 1 confirmed our hypothesis nearly perfectly. The only plausible explanation is related to the different composition of groups in Study 1 and 2, especially in terms of age. Treating another null result with caution, the limited scope of age in Study 2 could be the reason for not finding a significant increase of the number of L2 increases with increasing age.

#### 4.5.4 Conclusion

In conclusion, we found that seven minutes of exposure was not enough for children in the input group to outperform children in the control group in correctly rejecting non-words or correctly accepting real words that they heard for the first time in the LDT. Therefore, no evidence could be obtained for real learning from input in the child group. In contrast, adults in the input group did outperform adults in the control group in the ability to correctly reject CV\_nasal non-words. Since this structure is possible in the participants' native language and in the target language, but the specific instantiation of CV\_nasal is only illegal in the target language, we interpret this finding as being evidence for real learning from input. The fact that adults in the control group did not differ from adults in the experimental group in condition CV\_plosive sheds further light on the findings from Study 1. Results from Study 2 indicate that general inferencing mechanisms are sufficient to correctly reject CV\_plosives. Results from Study 1, therefore, are adequate to illustrate the development across the lifespan in dealing with novel L2 input. They are, however, not sufficiently detailed to unequivocally show learning from input. The implementation of two control groups in Study 2 compensated for this missing value. By means of introducing a consolidation period, Study 2 further analyzed the learning ability from the input after a consolidation period. The ability to correctly reject CV\_nasal non-words did not decline after a consolidation time of one week. Even though the ability did not improve at time two, we still consider this finding as consolidated learning, taking into account that participants received no more correct input before the second testing and received much 'bad' input through the course of the LDT. In sum, adults in the control and in the experimental groups differed in accuracy in condition CV\_nasal, while, in both groups, children differed from adults in terms of accuracy in all conditions except for CV\_nasal. The age effect is therefore attributed to variables unrelated to the exposure condition, such as generally increased experience with languages.

In sum, our findings are in line with studies challenging the Critical Period Hypothesis (cf. e.g. Cenoz, 2002; Dimroth & Haberzettl, 2012; Flege et al., 1999; Hakuta et al., 2003; Miralpeix, 2007; Muñoz, 2006; Neufeld, 1977; Snow & Hoefnagel-Höhle, 1978). Clearly, our results speak against findings that claim that adults are no longer able to break into an unknown language system, as proposed by Kuhl's Native Language Magnet Theory, for example (cf. also Colombo, 1982; Hyldenstam & Abrahamsson, 2003; 2004; Lenneberg, 1967; Weber-Fox & Neville, 1996).

## 5 Summary and General Conclusions

### 5.1 Summary of findings

The aim of this thesis was to investigate the very initial stages of natural and untrained implicit language learning. We were particularly interested in the influence of age as well as cognitive variables determined by age and the influence of multilingualism, represented here as the number of foreign languages a participant knew. Overall, we found that adults are capable of learning to distinguish sound regularities of a new language and generalize these to new items, and that they are able to retain this implicitly acquired knowledge over a retention period of one week. There was no age effect and no evidence for a declining capacity across the age span, neither for implicit learning, nor for the perception, distinction and generalization of foreign language phonotactics. Indeed, adults outperformed children, rather than vice versa.

Both studies used a paradigm whereby participants were first exposed to a seven-minute Weather Report (WR) in Mandarin Chinese and immediately thereafter completed a Lexical Decision Task (LDT) designed to test whether they had extracted phonotactic information about Mandarin Chinese implicitly in that brief time period. We analyzed accuracy and reaction time, as well as correlations between responses, age, cognitive variables (working memory, crystallized and fluid intelligence), and multilingualism. In Study 1, we tested three hypotheses (cf. Section 3.3.3):

- 1) Adults and children learn to distinguish sound regularities after 7 minutes in the following order of difficulty: (easiest) VCCC > VCC > CV\_plosive > CV\_nasal (most difficult);
- 2) Adults perform better than children in terms of accuracy (but not in terms of reaction time);
- 3) Each cognitive variable (Gc, Gf, WM and # L2s) correlates with different strengths with accuracy in the four syllable conditions: (strongest) VCCC > VCC > CV\_plosive > CV\_nasal (weakest correlation).

In Study 2, we tested three additional hypotheses (cf. Section 4.3.3):

- 1) The experimental groups perform significantly better than the control groups in the LDT; both in terms of accuracy and in terms of reactions times, but not in terms of working memory and the number of L2s;
- 2) Adults outperform children in those regards, except for reaction time;

- 3) The performance on the LDT stays stable or improves after one week.

In Study 1, age was treated as a continuous variable, starting from the age of 10/12 and 15/16, and nearly covering all ages between 20 to 86 years. Working memory, crystallized and fluid intelligence served as cognitive variables that were correlated with age as well as with the outcome of our experimental task. Multilingualism was also correlated with age and with accuracy. The three hypotheses were confirmed insofar as:

- 1) Participants learned to distinguish words from non-words after seven minutes of natural language input, except for CV\_nasal syllables.
- 2) The ability to generalize newly acquired knowledge from the input to new stimuli increased with increasing age for all syllable conditions except for CV\_nasal where the performance remained stable across the lifespan.
- 3) Crystallized intelligence and our measure for multilingualism positively and strongly correlated with response accuracy in the Lexical Decision Task, and so did fluid intelligence and working memory, but less strongly.

In Study 2, we used the same experimental paradigm, but compared two different age groups: 10-11-year-old children and 30-40-year old adults. To verify whether the effect from Study 1 was really attributable to the minimal exposure provided by the Weather Report or whether it was the result of a preconceived idea of what Mandarin Chinese sounds like, we also tested control groups that did not receive any input and therefore should not have had any formal knowledge of the language. Both the control groups and the experimental groups additionally provided the same measure for working memory and multilingualism as in Study 1. The three hypotheses were only partly confirmed:

- 1) We found a difference between the adult control and experimental groups for syllable condition CV\_nasal, but not for any other syllable condition, and neither did we find a significant difference in the performance between the child control and experimental groups for any of the syllable conditions. In contrast to Study 1, working memory did correlate with age and with hit rate in the experimental group, but not in the control group. Furthermore, multilingualism did not correlate in either group, neither with age nor with hit rate.
- 2) There was a significant age effect for all syllable conditions, except for condition CV\_nasal.
- 3) Adults also performed significantly better than children at time 1 and at time 2.

## 5.2 Theoretical Implications

### 5.2.1 Minimal exposure, implicit and statistical learning

The results of both our studies confirm results by Roberts et al. (2010) concerning the VCCC and VCC conditions. Unexpectedly, the ability to correctly reject CV\_plosive syllables also improved with age in both Study 1 and in Study 2. We interpreted this ability as a reflection of learning from input in Study 1, because the structure per se is possible in the participants' native language and in the target language. From Study 1, we concluded that the 'knowledge' (that the specific instantiation of CV\_plosive is illegal in the target language) had to be acquired from the input. However, adults in the experimental group in Study 2 did not differ significantly from adults in the control group in their ability to correctly reject CV\_plosive syllables. This suggests that general inferencing mechanisms or prior experience or knowledge were sufficient to correctly reject CV\_plosive syllables. Adults in both groups outperformed children, which points towards an advantage in general learning experience rather than implicit learning from input in condition CV\_plosive.

However, in condition CV\_nasal adults in the experimental group outperformed adults in the control group. This finding can be interpreted as evidence for implicit learning from input. The non-words in this syllable condition were supposed to be very difficult to reject, since the CVC-structure as such is possible in Chinese, even with CVC syllables ending in a nasal, but not with the specific nasal /m/ that appeared in CV\_nasal codas (/n/ and /ŋ/ are possible in the CVC coda position – cf. e.g. Lai, 2012). For the same reason, CV\_plosive was supposed to be difficult to be rejected as a non-word. However, since plosives are not at all possible in Mandarin Chinese syllable codas (Lee, 1976), and plosives therefore never appeared in the coda position in the Weather Report, CV\_plosive was supposed to be easier to classify than CV\_nasal (nasals /n/ and /ng/ appeared in coda positions in the WR). Moreover, CV\_plosive sounds more Germanic than CV\_nasal (cf. Lee, 1976). Consequently, our interpretation is that CV\_plosive syllables do not need input to be correctly rejected as non-words (cf. Kellerman, 1979, 1983; Kellerman & Sharwood Smith, 1986).

While many think that an L2 cannot be learned without negative evidence (cf. e.g. Chomsky, 1980; Gold, 1967; Pullum, 1996; White, 1991), others would agree that this is possible (cf. e.g. Bates & Elman, 1996; Bley-Vroman, 1991; Chater & Vitányi, 2007; MacWhinney, 2004; Zyzik, 2009), for example by using statistical pattern recognition. The ability to correctly reject CV\_nasal syllables also did not require the use of negative evidence (meaning an explicit indication that these syllables are not compatible with Mandarin Chinese phonotactics). But in this condition, participants were required to pay attention to input (which provided positive evidence). Therefore, the difference between implicit learning and making inferences arose in the condition where we can draw on input, namely in CV\_nasal.

Even though we must be careful in interpreting null results, it is interesting to note that adults in the input condition mastered the CV\_nasals significantly above chance whereas children did not. Children might have needed additional input to perform above chance for the CV\_nasal syllables. The length of exposure in our study was a result of the word count in the Weather Report, which again resulted from controlling the frequency of words clause-initially and -finally (cf. Section 3.2.2). Possibly, the effect of our study could be elicited with even less exposure, or could perhaps be enhanced with a slightly increased exposure (cf. Roberts et al., 2010). Adults in Roberts et al. (2010) significantly improved their performance on the CVC syllables after double exposure.

Together, these findings are consistent with accumulating evidence for an adult capacity to swiftly learn to process complex natural language material from novel L2 input even in the absence of conscious learning efforts and are comparable to implicit learning effects in children (cf. e.g. DeKeyser, 2003). It would be interesting to apply the same paradigm to an explicit learning situation, for example by instructing participants to pay attention to the Chinese language structure, in order to be prepared for the subsequent Lexical Decision Task. Most probably, a change in instructions and thereby a different focus of attention would affect the outcome. However, Seger (1994) doubts that it is possible to develop a task that either involves only implicit or only explicit learning (cf. p. 27). Cleeremans and Dienes (2008) were similarly convinced that a mixture of implicit and explicit learning is always involved in typical implicit learning situations (cf. p. 413).

### 5.2.2 Age, L1 influence and cognitive variables

The findings from the two studies imply that adults are still able to learn rapidly and implicitly, contrary to claims by, for example, DeKeyser (2003, 2012, 2013), and neurological arguments that claim the loss of plasticity in the adult brain to be evidence for adults' less successful L2 acquisition (cf. e.g. Kuhl, 2004; Long, 2005; Newport, 1990). Our results instead support findings suggesting that adults retain plasticity in the language system (e.g. Stein et al., 2006; Steinhauer et al., 2009) and therefore the potential for acquiring native-like proficiency in a new language (e.g. Abu-Rabia & Kehat, 2004; Friederici et al., 2002; Montrul & Slabakova, 2003; Neufeld, 1977; Reichle, 2010; Snow & Hoefnagel-Höhle, 1978; van Boxtel et al., 2005).

Changes in the brain following L2 learning are a sign of the accommodation of neural networks to new processing strategies even though little is known about the neurobiological correlates (cf. e.g. Mårtensson et al., 2012). Yet, increasing evidence supports the view that the same brain regions process the L2 as the L1 (cf. e.g. Chee, Tan, & Thiel, 1999; Friederici et al., 2002; Hernandez et al., 2000; Klein, Milner, Zatorre, Zhao, & Nikelski, 1999). Using fMRI in combination with the same experimental exposure as we did, Veroude et al. (2010) found stronger functional connectivity between areas implicated in phonological storage after the double-exposure to the Weather Report. These structural neurological adjustments were only present in participants who performed above chance in the Word Recognition Task, but not in participants who performed at chance level. Structural changes in the brain were also shown in longitudinal studies by McLaughlin et al. (2004) after 14 hours of L2 classroom instruction and by Osterhout et al. (2008) after 9 weeks of classroom instruction. Osterhout, McLaughlin, Pitkänen, Frenck-Mestre and Molinaro (2006) also showed changes in the brain's electrical activity after 14 hours of classroom instruction for the learning of L2 word form, after 60 hours for learning of word meaning, and after 140 hours event-related brain potentials became native-like. These results support our finding that learning takes place after minimal exposure. Again, possibly the effect would have been stronger with double exposure.

The explanation of our findings may be found in aspects of higher numbers of L2s, and crystallized or stored information, such as general knowledge, vocabulary and learned skills (cf. Cattell, 1987). This again might be related to more or less L1-influence on L2-processing, but that remains to be studied further. Since adults in our studies outperformed children, well-established L1 knowledge might actually

promote L2 word segmentation and not constrain L2 listening as suggested, for example, by Weber and Cutler (2006) or Finn and Hudson Kam (2008) (cf. also Cutler, 2001; Cutler & Otake, 1994; Flege & Wang, 1989). Dimroth and Haberzettl (2012) also argue alongside Weber and Cutler (2006) that L1 entrenchment constrains adult L2 learners' attention to certain aspects of the L2 input more than child L2 learners' attention, because children's L1 use is still less automatised (Dimroth & Haberzettl, p. 347). Dimroth and Haberzettl's argumentation is in line with MacWhinney's Unified Competition Model (2005), where repeated use of the L1 is emphasized as leading to ongoing L1 entrenchment, which is, however, expected to be strongest in output phonology and weakest in the area of lexicon (cf. p. 63).

There are different ideas about L1-influence in terms of an increased activation of competing sounds to the ability of L2-listening and -recognition (cf. also Altenberg, 2005b; Broersma, 2005; Broersma & Cutler, 2011; Weber & Broersma, 2012) and in terms of the perception of similar or dissimilar L2 sounds (cf. e.g. Best, 1995; Flege, 1993). While "new but not similar sounds in an L2 may be mastered" according to the Speech Learning Model (SLM) by Flege (1993, p. 1589), similarity of the L2 sound to the listener's L1 facilitates rather than hinders discrimination according to the Perceptual Assimilation Model (PAM) by Best (1995, p. 194; cf. also Snow & Hoefnagel-Höhle, 1978). Best et al. (2001) generally regarded a stable L1 to be support for the perception of foreign utterances, while Nenonen et al. (2005) stated that phoneme representations of the native language exert a strong [negative] influence on contrast detection, even in a well-learned second language. In the sense of PAM, better-established L1 phonotactic knowledge of adults in our studies might therefore promote better lexical decision performance, which would explain better performance on the L1-sounding non-words with increasing age.

We expected working memory capacity to increase into adulthood, especially through the maturation of the frontal cortex and the usage of the phonological loop (cf. Gathercole & Baddeley, 1993; Huttenlocher & Dabholkar, 1997), and decrease into old age (which was only represented in Study 1). The increase of WM with age was significant in the experimental group in Study 2. In these participants, working memory scores also predicted successful incidental learning, similarly to the study by Robinson (2005, 2010). The reason why we did not find a significant increase of working memory with age in Study 1 could be due to a different sample (i.e. in terms of the age span covered). In Study 2, we only tested participants up to the age

of 40, while our oldest participants in Study 1 were 86 years old and WM capacity is thought to start to decline in adulthood (Borella et al., 2008; Engle et al., 1999; Gathercole et al., 2004; Jenkins et al., 1999; Park et al., 2002; Park & Payer, 2006; Salthouse, 1994, 1996). On the other hand, the different correlation pattern could also be partly attributed to possible inter-rater variability. Even though we controlled the output of the examiners by recording the digit sequences (which otherwise would be produced by the examiner), this task still involved an uncontrollable amount of non-verbal interaction between the participants and the examiners during the execution of the task. No form of feedback, for example, was allowed before the termination of the task, but it is possible that some examiners showed more of a poker face than others and thereby encouraged or motivated participants more or less than other examiners. In Study 2 we had only 2 examiners instead of four (as in Study 1), which was probably an advantage. We did consider a more reliable task, for example the Operation Span Task (Turner & Engle, 1989) that is completely computerized for Study 1, but it would have been too time-consuming. Since we aimed at keeping the procedure as constant as possible, we also did not change the task for Study 2. Nevertheless, in Study 1, we saw a significant correlation of working memory scores and the performance in the syllable conditions VCCC and VCC. This might reflect the aforementioned effect of experience and age. Kormos and Sáfár (2008) also used the backward digit span task to measure working memory and found a correlation of WM with, amongst others, L2 listening ability.

Finally, the influence of ‘multilingualism’ may be related to the aforementioned cognitive advantage. Kavé et al. (2008), amongst others (e.g. Bialystok et al., 2004; Bialystok, Martin, et al., 2005; Craik et al., 2010; Perquin et al., 2013), found multilingualism to be a significant predictor of cognitive state, even more influential than age, education, place of birth or age at immigration. Using the Mini-Mental State Exam (MMSE Folstein, Folstein, & McHugh, 1975) the authors (Kavé et al., 2008) connected multilingualism to the theory of cognitive reserve that claims abilities such as education, occupation and intelligence to slow down cognitive aging processes in cognitively active elderly compared to others (e.g. Stern, 2002). Bialystok (2004) showed that bilinguals’ control processes are better than monolinguals’ and that this advance continues into (old) age. Perquin (2013) found the protection against cognitive impairment (without dementia) to be seven times higher in 65+-year-old participants that practiced between two to three languages instead of one or two. Results in Study 1 are in line with predictions of cognitive reserve, as multilingualism was significantly correlated with age and with accuracy

in the LDT. Although null results must be treated with caution, the relatively small number of participants per group, and especially the underrepresentation of elderly, could explain the absence of a significant correlation in Study 2.

### **5.2.3 Quasi-longitudinal perspective**

The implicit learning effect in condition CV\_nasal remained for adults in Study 2 after the consolidation time of one week. An interesting further question is whether this learning effect could have been enhanced by double exposure at time 1 and/or prior training. Auditory training, for example, has resulted in better L2 perception performance (cf. e.g. Bradlow et al., 1997; Hazan, Sennema, Iba, & Faulkner, 2005; Wang, Spence, Jongman, & Sereno, 1999). An additional interesting question is whether performance would improve under explicit instruction conditions, as suggested above, where participants' attention would be drawn to specific aspects of the input. This would trigger explicit learning mechanisms, which would differ from the current instructions that promoted implicit learning. One possibility to promote explicit learning could be to present pinyin-symbols on-screen during the Weather Report and during the Lexical Decision Task. Bassetti (2006) found that knowledge of Mandarin Chinese orthographic (pinyin) word forms influenced learners' memory of the phonological forms. In her study, native English-speaking learners of Mandarin counted the number of sounds more accurately if they saw the orthographic representation of vowels than if they did not. Possibly, learners of Mandarin phonotactics would be able to remember legal syllable structures better with the visual support. Recent findings have provided further evidence for the influence of orthographic forms on phonological form memory in L2 learners (cf. e.g. Detey & Nespoulous, 2008; Hayes-Harb, Nicol, & Barker, 2010; Showalter & Hayes-Harb, 2013). Showalter and Hayes-Harb (2013) suggest that learners' knowledge of a novel L2 orthographic feature supports the association of a novel L2 phonological feature with novel L2 words.

Furthermore, it would be interesting to examine effects of speaker or voice variability. In a study by Barcroft and Sommers (2005), for example, multiple-talker and -voice formats had a positive effect on L2 word learning (cf. also Bradlow et al., 1997; Hazan et al., 2005; Logan, Lively, & Pisoni, 1991; Wang et al., 1999). It would be possible, for example, to test participants after multiple exposures to the WR, once spoken by speaker A, and a second time by speaker B. Additional variation could be introduced by changing speakers or voice for each of the six weather charts. Speaker and voice variability as well as variability of phonotactics is

a trait of natural languages (cf. e.g. Dell et al., 2000). Artificial languages are less complex because they are often small, simplified and language tokens are repeated frequently or trained before discrimination tasks (cf. e.g. Fitch & Friederici, 2012). In order to acquire sensitivity to novel L2 phonotactics, L2 learners require experiences with L2 sound sequences in order to store them in memory. Since variability is often arbitrary, it seems to be important to expose learners to as much variability as possible in order to establish knowledge of new sound patterns (cf. e.g. Dell et al., 2000).

Adults in Study 2 were able to distinguish words from non-words and successfully classify CV\_nasal syllables without repetition or training of vocabularies or rules to guide their learning process, and without further input after the consolidation time of one week. Our findings suggest that implicit statistical learning mechanisms benefit from general experience, such as increased crystallized intelligence along the lifespan and/or possibly increased number of foreign languages across the lifespan.

In sum, our results support findings both in the artificial and statistical language learning literature (Friederici et al., 2002; Perruchet & Poulin-Charronnat, 2012; Saffran, Newport, et al., 1996; Saffran et al., 1997) and in studies on first exposure to natural language (Bisson et al., 2013; Carroll & Widjaja, 2013; Gullberg et al., 2012; Gullberg et al., 2010; Rast, 2008; Roberts et al., 2010; Shoemaker & Rast, 2013; Veroude et al., 2010), suggesting a malleable and continual implicit learning ability that can be triggered already after seven minutes of natural language input. What is more, contrary to popular belief and to the literature dealing with age effects in acquisition, this ability seems to improve or at least remain stable across the lifespan.

### **5.3 Comments on Methodology**

The goal of this thesis was to investigate the very earliest stages of L2 acquisition across the lifespan, in the absence of pre-existing knowledge about cognates and phonotactics to bootstrap and boost learning. The Mandarin Chinese Weather Report has proven most adequate for this purpose, given that it is constructed and therefore highly controlled, but most importantly, complex natural language is used in an audio-visual context that was chosen to be as authentic and brief as possible.

#### **5.3.1 Participant selection**

In Study 1, we collected data together with two other projects (Project ‘A’ and Project ‘C’), which on the one hand allowed us to reach more participants and collect more data, but on the other hand increased the total testing time for each participant, as each project had its specific experiment. The factor *time* made recruitment more difficult, since participants had to agree to a testing session of approximately three hours, which often proved difficult, especially for people between 30-60 years. Many situational factors had to be taken into account and therefore, our participant selection is not completely random, since ‘almost every sample has been one of convenience’ (DeKeyser, Alfi-Shabtay, & Ravid, 2010, p. 416). Moreover, this resulted in unbalanced distributions of male and female participants in certain age-bands. The most unbalanced distribution of men and women, for example, was present in the age band 30 to 39 (5 men, 14 women) and the age band 40 to 49 (3 men, 16 women), for the simple reason that women who stayed at home caring for children were more flexible time-wise.

In Study 2, time again proved to be a factor that influenced the representativeness of our sample. Testing time was significantly shorter (maximally 40 minutes at time 1); yet, proportional to the time for the way here/way there, the total financial incentive was relatively small, especially for the experimental adult group that had to come for testing twice and the second time for only ten minutes (children were tested at their school).

Two other factors that impeded random participant selection in both studies were *interest in L2s* and *money*. Since we could not force our sample to participate, we had to give some information about the study in order to awaken people’s interest. The possibility that we ended up with a sample that was particularly interested in

foreign languages is not to be neglected in the interpretation of the outcome. A second factor that influenced the representativeness of our sample was the financial incentive. People on a tight budget, possibly unstable employment situations, and/or homestay due to childcare, were probably more receptive to a small financial reimbursement than people with more money than time on their hands. Certainly, in the age band 20-29 (where students are typically represented), and in the age bands 30-39 and 40-49 (where available time is restrained in successful jobs), the probability that participants responded to our recruiting drive was influenced more strongly by the outlook of a financial compensation than in age bands where time is less restrained. Possibly, these two factors, interest in L2s and money, cancel each other out, insofar as we might have both kinds of people in our sample - the ones interested in language learning, and the ones mostly interested in our compensation.

### **5.3.2 LDT paradigm**

The LDT paradigm is often used for semantic priming experiments (cf. e.g. Antos, 1979), where mostly reaction times are analyzed and accuracy is only secondarily of interest. Since Roberts et al. (2010) did not find any significant effect for reaction time, we did not expect to find a significant effect either. However, we did find significant effects of age, syllable condition and reaction time in both studies. In Study 1, reaction time was affected by age in the critical syllable conditions CV\_nasal and CV\_plosive where reactions times increased across the lifespan. In Study 2, there was no significant age effect for reaction time, most probably because the higher age spectrum was not covered. However, there was a significant effect of reaction time and group in Study 2, which was again carried by the critical conditions CV\_nasal. Adults in the experimental group performed significantly faster in this condition compared to adults in the control group. This suggests that adults in the experimental group performed faster because of a learning effect. The development of reaction time in Study 1 would therefore support Ratcliff's and others' (cf. also Hale & Myerson, 1995; Myerson, Ferraro, Hale, & Lima, 1992; Ratcliff, Thapar, Gomez, & McKoon, 2004) claim that older people perform more slowly but more accurately, while Study 2 showed the effect of exposure (in condition CV\_nasal) even for reaction time.

In our Lexical Decision Task, 50 percent of the stimuli were real words, while 50 percent were non-words. As a result of this ratio of valid versus invalid trials, participants received more and more 'false' input towards the end of the LDT. This could have lead to a decline in the learning effect after the Weather Report.

MacWhinney (2004), for example, emphasized that “the provision of good quality positive evidence” (p. 911) is the most important to language development. On the other hand, participants listened to Chinese words ‘in isolation’ for the first time during the LDT. Based on Peña et al.’s (2002) and Seidenberg et al.’s (2002) findings that the introduction of pauses in the speech stream increased segmentation performance, we could argue that learning could have still taken place during the LDT. Furthermore, there was no deterioration of accuracy at time 2. If participants had performed significantly worse at time 2, it would have been difficult to explain with certainty what caused the effect. Explanations could reach from a mere temporary learning effect at time one to other effects, such as fatigue, boredom, or forgetfulness of which button was which (Chinese/ Not Chinese). On the other hand, participants could have performed significantly better at time two and there would be no way to subtract the effect of ‘false’ input out of the equation. In a future experiment, it would be interesting to examine the effect of the ratio of words and non-words in the LDT. In addition, it would be worthwhile to consider an alternative or additional experimental paradigm (for ideas see Outlook, Section 5.5).

### **5.3.3 Language background & proficiency**

Language background and language proficiency will always be difficult to control for, since “no two speakers have the same language, because no two speakers have the same experience of language” (Hudson, 1996, p. 11). We were frequently confronted with the question of how we controlled ‘the very initial state’, a first contact situation with Mandarin Chinese in our participants. Clearly, it was a selection criterion that participants did not know any Chinese, yet we cannot exclude the possibility that participants have been to a Chinese restaurant or seen a Chinese movie, since this possibility constitutes the natural context our participants live in and might therefore be seen as part of the initial state of natural language learning in the modern world. Kellerman (1983) defined psychotypology as the learner’s perceived distance of the L1 and the L2. Our results, like the ones found by Rast (2008, p. 231), suggest that “prior linguistic knowledge and metalinguistic strategies make up the learners’ psychotypologies”. Each learner’s psychotypology is thought to be individual in the same way that “no two speakers speak any given language in exactly the same way” (ibid.). In consequence, we might question the existence of a common ‘initial state’ or starting point for L2 acquisition (ibid.). In agreement with Rast (2008), we emphasize the need to consider learner variability in models of SLA.

Moreover, the problem with self-evaluation is also that ‘knowing’ or ‘not knowing’ can be regarded very subjectively, as well as ‘having had any contact with’ a language. Most people would probably not remember a Chinese restaurant as ‘having had contact with Chinese’, while they might pause to think on the matter after they travelled to China. However, most people would probably not enlist a language as part of their foreign language repertoire as long as they have not had the opportunity or motivation to practice the language (cf. e.g. Ellis, 2002).

Self-assessment in terms of proficiency is further problematic as different personality traits influence whether people either over- or underestimate their skills (North & Jones, 2009). Language teachers experience people with less meta-linguistic knowledge about their language skills as being more prone to overestimate their proficiency. Students have been found to judge their proficiency much more strictly after a completed language course than at the beginning - sometimes they gave themselves even less points after the course than before (personal conversation with language teacher in October 2011). That might be a reason why proficiency scores were rarely used in explaining variance obtained in the dependent variables of a study, as mentioned by Hulstijn (2012) in his review on language proficiency in the study of bilingualism. In our study, we only applied rough proficiency criteria to recruit participants, but did not exclude them on the basis of the proficiency measures they provided in the second (language background) questionnaire. We did not try to explain the outcome variance of the Lexical Decision Task with the proficiency measures of the different foreign languages the participants provided, but this is something that would be interesting to look at in a possible future study, as well as interactions with motivation and frequency of L2 use or meta-linguistic knowledge (compare Herdina & Jessner, 2000). Hulstijn (2012) advised caution in comparing proficiency scores of different languages with the same label, since the number and nature of language-specific linguistic elements differ (cf. p. 427). We used the CEFR proficiency levels developed by the Council of Europe (2011), even though Hulstijn argued in a similar vein “the six levels cannot be regarded as forming an unidimensional ladder of language development because the higher levels (B2, C1 and C2) can only be attained by people with higher intellectual capacities” (Hulstijn, 2012, p. 429). We asked participants to rate their listening, reading and writing capacity in the languages that they know, and since we did not test their skills, they readily assigned themselves B2- and C1-levels. It would be fruitful to analyze correlations of participants’ results on the English-test we

administered with their self-assigned proficiency-level for English. Strong correlations would support the belief in self-evaluation as a valuable tool.

Proficiency plays a significant role for the understanding of language acquisition and is seen to be more important in the evaluation of foreign language skills than, for example, age of acquisition (Abutalebi, Cappa, & Perani, 2001). Van Hell and Tanner (2012) suggested that in L2 syntactic processing, the proficiency variable may interact with cognitive variables such as working memory, attention, and inhibition. In a discussion of empirical studies on the relationship between L1 and L2 lexical processing, they showed that higher L2 proficiency was associated with increased attentional control and a greater ability to ignore irrelevant or inappropriate information.

## 5.4 Practical Implications

Obviously, the perceptual learning examined in the LDT is very different from successful L2 learning. Yet, the natural L2 learning context outside the classroom challenges learners exactly with the situation mimicked by our experimental paradigm: How to perceive and process unknown speech information with no help to break into the new language. It is therefore important to understand how learners go about this task in order to understand and/or improve how we could better guide learners in the L2 acquisition process. It is possible that age effects are more visible in production, typically examined in studies of ultimate attainment and nativelikeness, than in comprehension and perception studies. Our results suggest a constant or even increasing capacity along the lifespan to perceive and generalize newly acquired phonotactic knowledge. It remains an important challenge for future research to examine the potential relationship between production and comprehension and possible differing age effects on nativelikeness across these domains.

A caveat, however, is that ‘nativelikeness’ itself is not an unproblematic notion when considering speakers with varying and multilingual language experiences. A monolingual is not comparable to a bi- or multilingual. There is now plenty of psycholinguistic evidence to suggest that a bi- or multilingual brain simultaneously uses the L1 and the L2(s) while processing any foreign language - a task that entails additional executive control- and subcortical processes and that is therefore hardly comparable to processing only one language (e.g. Abutalebi, 2008; Abutalebi & Green, 2007; Friederici et al., 2002; Grosjean, 1989; Herdina & Jessner, 2002; Kroll, 2008). Usage-based approaches to language acquisition (e.g. Ellis, 2006; Ortega, 2013) also state that ”an individual’s creative linguistic competence emerges from the collaboration of the memories of all the utterances in their entire history of language use and from the frequency-biased abstraction of regularities within them” (Ellis, 2006, p. 2). Multilingualism and creativity seem to go hand in hand and it has been shown that creativity and procedural knowledge are better developed in multilingual than in monolingual children (Beardmore, 2008).

This, in turn, means that multilingual experiences will affect the whole (perceptual) learning system, making a monolingual native standard highly problematic. Such a view has potential practical implications, for example for instructed language

learning and teaching. DeKeyser (2003) stated that teaching methods should be adapted to the circumstances instead of blindly setting the age of onset to as early as possible. Schools can often not provide conditions for implicit learning, because “[...] time is limited and learning highly structured [...]” (DeKeyser, 2003, pp. 335, 336). In a related vein, Muñoz (2011) emphasized that sufficient intensity is needed for implicit learning to take place, both in terms of amount of input and intensive interactions with well-trained teachers and age-appropriate materials. Muñoz (2006) provided some support for the long-standing notion that adult learners have an advantage at the initial rate of learning, while child learners have an advantage at implicit learning (cf. Krashen et al., 1979). Comparing elementary and secondary students, Cenoz (2002) found better pronunciation scores in the early onset group, but better overall performance in oral proficiency, reading, writing and grammar in the late onset group. Muñoz (2006) specified that child learners would not outperform adults in the long run if similar exposure and instruction conditions were provided, since young learners need much more input in order to learn implicitly. Nikolov and Mihaljevic-Djigunovic (2011) have similarly pointed out how complex the relationships are between the (early) language learning capacity and the development of cognitive and affective skills, and how these interactions can provide us with insight into the multi-competence (cf. Cook, 1992) that emerges from the very beginning of foreign language learning. The current study has highlighted how remarkably little experience can make a difference allowing for highly abstract types of knowledge to emerge. Both our studies support Marinova-Todd’s (2000) finding that researchers comparing age effects in children and adults’ language acquisition often misinterpret differences in learning situations as age differences. As an example, Hakuta (2003) and Flege (1999) showed that age of learning effects disappeared once the level of education was controlled.

## 5.5 Outlook

This experimental paradigm has so far been run on Dutch and Swiss-German speaking adults and children. In a future study, it would be worth investigating how participants of non-Germanic language groups would solve our experimental task. Syllables have widely different structures across languages and in order to find out how non-native syllables are processed, we need to compare how participants of different L1s process the L2 syllables under investigation (cf. Cutler et al., 1986).

Another interesting question is whether performance would improve under explicit instruction conditions where participants' attention would be drawn to specific aspects of the input. Notice, however, that the implicit learning effect would be compromised in the context of explicit instruction. Toro, Sinnott and Soto-Faraco (2005), for example, showed that the implicit learning effect of segmentation did not occur in the condition where participants had to monitor the syllable stream for pitch changes. The purpose of the current instructions was to promote implicit learning. One possibility to promote explicit learning could be by presenting pinyin-symbols on screen during the Weather Report and during the Lexical Decision Task. VCCC syllables served as control syllables to make sure participants stayed on task. Given the near-ceiling effect for this syllable condition, we would consider exchanging the syllables by more CVC syllables and examine how participants would treat CVC syllables with illegal consonants in the syllable-onset compared to our CVC syllables with illegal consonants in the syllable-offset. The hypothesis would be that participants would need at least double exposure to the Weather Report before they could reject CVC syllables with illegal consonants in the onset significantly above chance level (cf. Roberts et al., 2010). In artificial language learning studies, training and or prior knowledge helped segmentation (cf. e.g. Lew-Williams & Saffran, 2012).

Furthermore, in a future experiment, we would like to investigate possible neurological changes after the exposure to the WR, such as demonstrated by Veroude et al. (2010). It would be interesting to replicate their findings by dividing participants into a learner and a non-learner group according to their performance in the LDT. Furthermore, it would be potentially rewarding to examine a child group with functional imaging as well. For both tasks (LDT and WRT), a parallel EEG examination would be fruitful in order to examine reaction time more closely.

Lexical decisions have been examined in a positron emission tomography (PET) study by Specht et al. (2003). They found that the rejection of non-words mainly required phonological discrimination processes while the discrimination of real words more strongly required lexical access. PET-studies are fascinating for their high spatial precision, unfortunately however, they are very expensive.

Importantly, even though the present results suggest implicit learning after minimal exposure and consistent learning effects after one week of consolidation time, the question remains as to how long this learning effect will remain. Especially since children were shown to outperform adults at later stages of learning, it remains a question for future research to examine for how long adults would be able to process L2 input more efficiently than children and where possible limitations would arise. The focus of Study 1 and Study 2 was on perception and generalization only, because “phonotactic constraints are first encountered and acquired through listening to language” (Kittredge & Dell, 2011, p. 2679). Successful second language learning, however, is typically measured in production. Dell et al. (2000) have shown that the language production system quickly adapts to phonotactic experience (cf. also Kittredge & Dell, 2011). An important question would therefore be how much exposure to natural language would be required for participants to produce correct L2 output and how perception and production skills interact in acquisition.

## 5.6 Conclusions

Results in both Study 1 and Study 2 challenge the notion of a critical period for foreign language learning in terms of the perception and generalization ability after minimal L2 natural input. Our findings are in line with results of previous studies using the same experimental paradigm (Gullberg et al., 2012; Roberts et al., 2010; Veroude et al., 2010). While all studies found implicit learning in adults after minimal exposure, our data extended the findings by adding a lifespan perspective of the development of this capacity, especially with respect to the correlation of cognitive variables and age. Furthermore, by examining the learning ability after one week, we added a short-term longitudinal perspective to our finding. In this respect, our data enabled some conclusions to be drawn on consolidation processes and the differences between children and adults. The fact that performance did not decrease after one week can be seen as evidence of memory processes (cf. Tamminen & Gaskell, 2006). However, to study changes in the rate of learning requires carefully designed longitudinal studies, which remains a major challenge for future studies.

In conclusion, the present studies have allowed us to investigate the human capacity to detect statistical regularities and generalize new abstract knowledge about an unknown language in the very first minutes of exposure. The Chinese Weather Report provided rich input to promote and examine implicit learning across the lifespan. Our results support findings both in the artificial language learning literature and in studies on first exposure to natural language. Overall, our findings are in line with usage-based approaches to language learning, suggesting a powerful mechanism to acquire and generalize information after minimal contact with a new language, even without pre-existing knowledge or pedagogical help to break into the new language system. In particular, this thesis provided some evidence for the claim that the capacity of detecting regularities in complex natural language input seems to benefit from more experience, both in terms of age and in terms of language skills.



## References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- 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.
- Abu-Rabia, S., & Kehat, S. (2004). The Critical Period for second language Pronunciation: Is there such a thing? Ten case studies of late starters who attained a native-like Hebrew accent. *Educational Psychology, 24*(1), 77-98.
- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol (Amst), 128*(3), 466-478.
- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition, 4*, 179-190.
- Abutalebi, J., & Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242-275.
- Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging, 25*(3), 736-740.
- Allen, D. (1992). *20-item multiple choice grammar test: first 20 items from the Oxford Placement Test* (2 ed.). Oxford: Oxford University Press.
- Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment, 25*(2), 1-7.
- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 80*(2), 606-621.
- Alloway, T. P., & Alloway, R. G. (2013). Working memory across the lifespan: a cross-sectional approach. *Journal of Cognitive Psychology, 25*(1), 84-93.
- Altenberg, E. P. (2005a). The judgment, perception, and production of consonant clusters in a second language. *International Review of Applied Linguistics in Language Teaching, 43*(1), 53-80.
- Altenberg, E. P. (2005b). The perception of word boundaries in a second language. *Second Language Research, 21*(4), 325-358.
- Anderson, V. A., Anderson, P., Northan, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology, 20*(1), 385-406.
- Antos, S. J. (1979). Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance, 5*(3), 527-545.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. *The psychology of learning and motivation, 2*, 89-195.
- Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F., & Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiology of learning and memory, 89*(1), 76-80.
- Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

- Baddeley, A. D. (2000). Working memory and language processing. In B. E. Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting* (pp. 1-16). Amsterdam: Benjamins.
- Baddeley, A. D., & Hitch, G. J. L. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- Baddeley, A., Gathercole, S., & Papagno, C. . (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.
- Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: on the dynamics between growth and decline. *Developmental Psychology*, *23*(5), 611-626.
- Barcroft, J. , & Sommers, M. S. . (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*(27), 387–414.
- Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign Language. *Written Language and Literacy*, *9*(1), 95-114.
- Bates, E., & Elman, J. (1996). Learning rediscovered. *Science*, *274*(5294), 1849-1850.
- Beardmore, H. B. (2008). Multilingualism, cognition and creativity. *International CLIL Research Journal*, *1*(1), 4-19.
- Bertrán, A. P. (1999). Prosodic typology: on the dichotomy between stress-timed and syllable-timed languages. *Language Design*, *2*, 103-130.
- Best, C., McRoberts, G. W. , & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Acoustical Society of America*, *109*(1), 775-794.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.
- Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Canadian Journal of Experimental Psychology*, *65*(4), 229-235.
- Bialystok, E., Craik, F. I., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: evidence from MEG. *Neuroimage*, *24*(1), 40-49.
- Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging*, *19*(2), 290-303.
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *J Exp Psychol Learn Mem Cogn*, *34*(4), 859-873.
- Bialystok, E., & Hakuta, K. (1994). *In other words: the science and psychology of second-language acquisition*. New York: Basic Books.
- Bialystok, E., & Hakuta, K. (1999). Confounded age: linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161-181). Mahwah, NJ: Lawrence Erlbaum.
- Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Dev Sci*, *7*(3), 325-339.
- Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: the rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103-119.
- Binet, A., & Simon, T. (1905). Application des méthodes nouvelles au diagnostic du niveau intellectuel chez les enfants normaux et anormaux d'hospice et d'école primaire. *L'Année Psychologique*, *11*, 245-336.

- Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language and Cognitive Processes*, 68, 706-755.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. de Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 88-108). New York: Oxford University Press.
- Birdsong, D. (2006). Age and second language acquisition and processing: a selective overview. *Language Learning*, 56(1), 9-49.
- Birdsong, D., & Gertken, L. M. (2013). In faint praise of folly. A critical review of native/non-native speaker comparisons, with examples from examples of native and bilingual processing of French complex syntax. *Language, Interaction and Acquisition*, 4(2), 107-133.
- Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, 44(2), 235-249.
- Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2013). Incidental acquisition of foreign language vocabulary through brief multi-modal exposure. *PLOSone*, 8(4), 1-7.
- Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. E. Rutherford & S. M. Sharwood (Eds.), *Grammar and second language teaching: a book of readings*. Boston, MA: Heinle and Heinle Publishers.
- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge: Cambridge University Press.
- Bley-Vroman, R. (1991). The logical problem of foreign language learning. *Linguistic Analysis*, 20, 3-49.
- Bongaerts, T., Planken, B., & Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 30-50). Clevedon: Multilingual Matters.
- Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychol (Amst)*, 128, 33-44.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101, 2299-2310.
- Broersma, M. (2005). *Phonetic and lexical processing in a second language*. (Unpublished Ph.D.), Radboud University, Nijmegen.
- Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, 36(1), 22-34.
- Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Q J Exp Psychol (Hove)*, 64(1), 74-95.
- Bugg, J. M., Zook, N. A., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2006). Age differences in fluid intelligence: contributions of general slowing and frontal decline. *Brain and cognition*, 62(1), 9-61.
- Bundesamt für Statistik (BFS) (2010). *Maturitäten und Übertritte an Hochschulen 2009*. Neuchâtel.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Carroll, S. E. (1999). Putting 'input' in its proper place. *Second Language Research*, 15, 337-388.
- Carroll, S. E. (2000). *Input and evidence: The raw material of second language acquisition*. Philadelphia: John Benjamins.
- Carroll, S. E. (2004). Learning how to 'hear words' in the L2 speech stream. *Transactions of the Philological Society*, 102(2), 227-254.

- Carroll, S. E., & Widjaja, E. (2013). Learning exponents of number on first exposure to an L2. *Second Language Research*, 29(2), 201-229.
- Carroll, Susanne E. (2002). Induction in a modular learner. *Second Language Research*, 18(3), 224-249.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of educational psychology*, 54(1), 1-22.
- Cattell, R. B. (1987). *Intelligence: its structure, growth, and action*. . New York: Elsevier Science Pub. Co.
- Cavanaugh, J. C., & Blanchard-Fields, F. (2006). *Adult development and aging* (5 ed.). Belmont, CA: Wadsworth Publishing/Thomson Learning.
- Cenoz, J. (2002). Age differences in foreign language learning. *ITL Review of Applied Linguistics*, 135-136, 125-142.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87(2), B69-B77.
- Chan, R. C. K., Shum, D., Touloupoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, 23(2), 201-216.
- Chater, N., & Vitányi, P. (2007). 'Ideal learning' of natural language: positive results about learning from positive evidence. *Journal of Mathematical psychology*, 51(3), 135-163.
- Chee, M. W. L., Tan, E. W. L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *Journal of Neuroscience*, 19, 3050-3056.
- Chen, C. Y. (1991). Shengdiao de zhuanbian yu kuosan: Taipei butong nianlingqun de quyang [The tonal change and diffusion: From the sampling of different age groups in Taipei]. *Journal of Chinese Language Teachers Association*, 16, 69-99.
- Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, 32, 867-873.
- Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes*, 13(2/3), 221-268.
- Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: evidence from event-related brain potentials *Language and Cognitive Processes*, 27, 231-256.
- Christiansen, M. H., Onnis, L., & Hockema, S. A. (2009). The secret is in the sound: from unsegmented speech to lexical categories. *Dev Sci*, 12(3), 388-395.
- Cleeremans, A., & Dienes, Z. (2008). Computational models of implicit learning *Cambridge handbook of computational cognitive modeling* (pp. 396-421).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.): Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3 ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Colombo, J. (1982). The critical period concept: research, methodological, and theoretical issues. *Psychol Bull*, 91, 260-275.
- Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition*, 114(3), 356-371.
- Cook, V. (1992). Evidence for multi-competence. *Language Learning*, 44(4), 557-591.
- Coppieters, R. (1987). Competence differences between natives and near-native speakers. *Language and Cognitive Processes*, 63, 544-573.

- Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 4, 161-170.
- Council of Europe, CEFR. (2011). *Common European Framework of Reference for Languages: learning, teaching, assessment.*: Cambridge University Press.
- Cowan, N. (1995). *Attention and memory: an integrated framework*. Oxford: Oxford University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169, 323-338.
- Cowan, N., & Alloway, T. P. . (2008). The development of working memory. In N. Cowan (Ed.), *Development of memory in childhood* (2 ed., pp. 303-342). Hove, England: Psychology Press.
- Craik, F. I. M., Bialystok, E., & Freedman, M. (2010). Delaying the onset of alzheimer disease. *Neurology*, 75, 1726-1729.
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology*, 30(5), 1119-1130.
- Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, 5(1), 1-23.
- Cutler, A. , & Otake, T. (1994). Mora or phoneme – further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824–844.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385–400.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1989). Limit on bilingualism. *Nature*, 340, 229-230.
- Daneman, M., & Carpenter, P. A. (1980). Individual difference in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, 21(4), 803-820.
- De Diego-Balaguer, R., & Lopez-Barroso, D. (2010). Cognitive and neural mechanisms sustaining rule learning from speech. *Language Learning*, 60(2), 151-187.
- De Luca, C. R., & Leventer, R. J. (2008). Developmental trajectories of executive functions across the lifespan. In P. Anderson & V. Anderson (Eds.), *Executive functions and the frontal lobes: a lifespan perspective* (pp. 3-21). Washington, DC: Taylor & Francis.
- Dehaene, S. (1999). Fitting two languages into one brain. *Brain*, 122(12), 2207-2208.
- DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. . (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, 31, 413–438.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 493-533.
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Oxford: Blackwells.
- DeKeyser, R. M. (2012). Age effects in second language learning. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 442-460). New York: Routledge.
- DeKeyser, R. M. (2013). Age effects in second language learning: stepping stones toward better understanding. *Language Learning*, 63(1), 52-67.

- Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). *The comparative fallacy in L2 processing research*. Paper presented at the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006), Somerville, MA.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors phonotactic constraints implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1355-1367.
- Detey, S., & Nespoulous, J.-L. (2008). Can orthography influence second language syllabic segmentation? *Lingua*, 118(1), 66-81.
- Diamond, M. C. (1996). The brain... Use it or lose it. *Mindshift Connection*, 1(1), 1.
- Diehl, R. L., & Lindblom, B. (2004). Explaining the structure of feature and phoneme inventories. In S. Greenberg, W. Ainsworth, A. Popper & R. Fay (Eds.), *Speech processing in the auditory system* (pp. 101-162). New York, NJ: Springer.
- Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep medicine reviews*, 13(5), 309-321.
- Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioural and Brain Sciences*, 22, 735-808.
- Dimroth, C., & Haberzettl, S. (2012). The older the better, or more is more: language acquisition in childhood. In M. Watorek, S. Benazzo & M. Hickmann (Eds.), *Comparative perspectives to language acquisition: a tribute to Clive Perdue* (pp. 324-349). Clevedon: Multilingual Matters.
- Ellis, N. C. (1994). Implicit and explicit language learning: an overview. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1-32). San Diego, CA: Academic Press.
- Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188.
- Ellis, N. C. (2006). Cognitive perspectives on SLA. *AILA Review*, 19, 100-121.
- Ellis, N. C. (2012). Frequency-based accounts of SLA. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*. New York: Routledge.
- Ellis, N. C., & Sagarra, N. (2010). Learned attention effects in L2 temporal reference: the first hour and the next eight semesters. *Language Learning*, 60(2), 85-108.
- Elman, J. L. (1990). Finding structure in time. *Cogn Sci*, 14, 179-211.
- Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48, 71-99.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cogn Sci*, 33, 1-36.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309-331.
- Escudero, P. (2005). *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. (Doctoral dissertation), Utrecht University, LOT Dissertation Series 113.
- Espy, K. A. (2004). Using developmental, cognitive, and neuroscience approaches to understand executive functions in preschool children. *Developmental Neuropsychology*, 26(1), 379-384.
- Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, 425(6958), 614-616.
- Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, 108(2), 477-499.

- Fischer, S., Drosopoulos, S., Tsun, J., & Born, J. (2006). Implicit learning – explicit knowing: a role for sleep in memory system interaction. *Journal of Cognitive Neuroscience*, 18(3), 311–319.
- Fischer, S., Wilhelm, I., & Born, J. (2007). Developmental differences in sleep's role for implicit off-line learning: comparing children with adults. *Journal of Cognitive Neuroscience*, 19(2), 214–227.
- Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: an overview. *Philosophical Transactions of The Royal Society*, 367, 1933–1955.
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, 93(3), 1589–1608.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: differences and similarities* (pp. 319–355). Berlin: Mouton de Gruyter.
- Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Bristol, England: Multilingual Matters.
- Flege, J. E., Frieda, A. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25, 169–186.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134.
- Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t-/d/ contrast. *Journal of Phonetics*, 17, 299–315.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78–104.
- Folia, V., Uddén, J., de Vries, M. H., Forkstam, C., & Petersson, K. M. (2010). Artificial language learning in adults and children. *Language Learning*, 60(2), 188–220.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *Psychiatric Research*, 12, 189–198.
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc Natl Acad Sci U S A*, 99(1), 529–534.
- Frisch, S. F., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (2001). *Emergent phonotactic generalizations in English and Arabic*. Paper presented at the Symposium on Frequency Effects and Emergent Grammar, Amsterdam.
- Fromkin, V. (1974). The development of language in Genie: a case of language acquisition beyond the 'critical period'. *Brain Lang*, 1(1), 81–107.
- Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *The Journal of Neuroscience*, 29(27), 8726–8733.
- Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart and Winston. Inc., 1970.
- Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory*(13), 3.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377.
- Gass, S. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Erlbaum.
- Gathercole, S. E. (2008). Working memory. In H. L. Roediger (Ed.), *Cognitive Psychology of Memory* (Vol. 2, pp. 33–52). Oxford: Elsevier.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Mahwah, NJ: Erlbaum.

- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177-190.
- Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 34-54.
- Gibson, E., & Gibson, J. J. (1955). Perceptual learning: differentiation or enrichment? *Psychological Review, 62*(32-41).
- Gold, E. M. (1967). Language identification in the limit. *Information and Control, 16*, 447-474.
- Golestani, N., Alario, F. X., Meriaux, S., Le Bihan, D., Dehaene, S., & Pallier, C. (2006). Syntax production in bilinguals. *Neuropsychologia, 44*, 1029-1040.
- Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*(3), 640-665.
- Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology, 16*, 562-576.
- Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew-English bilinguals. *Journal of Neurolinguistics, 17*, 31-52.
- Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word, 20*, 157-177.
- Gregoire, J., & Van der Linden, M. . (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition, 4*(2), 140-149.
- Grosjean, F. . (1989). Neurolinguistics, beware! The bilingual is not two monolinguals in one person. *Brain Lang, 36*, 3-15.
- Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics in Language Teaching, 50*, 239-276.
- 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*(2), 5-24.
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: a test of the critical period hypothesis for second language acquisition. *Psychological Science, 14*, 31-38.
- Hale, S., & Myerson, J. (1995). Fifty years older, fifty percent slower? Meta-analytic regression models and semantic context effects. *Aging and Cognition, 2*, 132-145.
- Hale, S., Rose, N. S., Myerson, J., Strube, M. J., Sommers, M., Tye-Murray, N., & Spehar, B. (2011). The structure of working memory abilities across the adult life span. *Psychology and Aging, 26*(1), 92-110.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-226). New York: Academic Press.
- Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research, 23*(1), 65-94.
- Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech, 53*, 367-381.
- Hayward, K. (2000). *Experimental phonetics: an introduction*. Harlow: Longman.
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication, 47*, 360-378.

- Herdina, P., & Jessner, U. (2000). The dynamics of third language acquisition. Bilingual education and bilingualism. In J. Cenoz & U. Jessner (Eds.), *English in Europe: the acquisition of a third language* (pp. 84-98). Clevedon, England: Multilingual Matters.
- Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: perspectives of change in psycholinguistics*. Clevedon: Multilingual Matters.
- Hernandez, A. E., & Kohnert, K. J. (1999). Aging and language switching in bilinguals. *Aging, Neuropsychology, and Cognition*, 6, 69-83.
- Hernandez, A. E., Martinez, A., & Kohnert, K. (2000). In search of the language switch: an fMRI study of picture naming in Spanish-English bilinguals. *Brain Lang*, 73, 421-431.
- Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, 9, 220-225.
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychol (Amst)*, 26, 107-129.
- Hudson, R. A. (1996). *Sociolinguistics* (2 ed.). New York: Cambridge University Press.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford: Blackwells.
- Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition*, 27, 129-140.
- Hulstijn, J. H. (2007). Psycholinguistic perspectives on language and its acquisition. In J. Cummins & C. Davison (Eds.), *International handbook of english language teaching* (pp. 701-713). Norwell, MA: Springer.
- Hulstijn, Jan H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, 15(02), 422-433.
- Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *The Journal of Comparative Neurology*, 387, 167-178.
- Hyashizaki, Y. (2004). Structural plasticity in the bilingual brain. *Nature*, 431, 757.
- Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-588). Amsterdam: John Benjamins.
- Indefrey, P. (2006). It is time to work toward explicit processing models for native and second language speakers. *Applied Psycholinguistics*, 27, 66-69.
- Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, 16, 73-98.
- Jacobs, J. (1887). Experiments on "prehension". *Mind*, 12, 75-79.
- Jakobovits, L. A. (1970). *Foreign language learning: a psycholinguistic analysis of the issues*. Rowley, MA: Newbury house.
- Jakobson, R. (1969). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Frankfurt am Main: Suhrkamp Verlag.
- Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: age-related differences in implicit sequence learning across the human lifespan. *Dev Sci*, 15(4), 496-505.
- Jenkins, L., Myerson, J., Hale, S., & Fry, A. F. (1999). Individual and developmental differences in working memory across the life span. *Psychonomic Bulletin and Review*, 6, 28-40.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn Psychol*, 21(1), 60-99.

- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review*, 9(4), 637-671.
- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. J. (2010). Implicit learning as an ability. *Cognition*, 116, 321-340.
- Kavé, G., Eyal, N., Shorek, A., & Cohen-Mansfield, J. (2008). Multilingualism and cognitive state in the oldest old. *Psychology and Aging*, 23(1), 70-78.
- Kellerman, E. (1979). Transfer and non-transfer: where we are now. *Studies in Second Language Acquisition*, 2(1).
- Kellerman, E. (1983). Now you see it, now you don't. *Language transfer in language learning*, 54(12), 112-134.
- Kellerman, E., & Sharwood Smith, M. S. (1986). *Crosslinguistic influence in second language acquisition*: Pergamon Institute of English.
- Kittredge, A. K., & Dell, G. S. (2011). *Implicit learning of phonotactic constraints: transfer from perception to production*. Paper presented at the 33rd Annual Conference of the Cognitive Science Society. , Austin.
- Klein, D., Milner, B., Zatorre, R. J., Zhao, V., & Nikelski, J. (1999). Cerebral organization in bilinguals: a PET study of Chinese-English verb generation. *NeuroReport*, 10, 2841-2846.
- Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.
- Klein, W. (1998). The contribution of second language acquisition research. *Language Learning*, 48, 527-550.
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11, 261-271.
- Krashen, S. (1973). Lateralization, language learning, and the critical period. *Language Learning*, 23, 63-74.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. London: Pergamon Press.
- Krashen, S., Long, M. H., & Scarcella, R. (1979). Age, rate and eventual attainment in second language acquisition. In S. D. Krashen, R. C. Scarcella & M. H. Long (Eds.), *Child-adult differences in second language acquisition* (Vol. 9, pp. 161-172). Rowley, Mass: Newbury House Publishers.
- Kroll, J. F. (2008). Juggling two languages in one mind. *Psychological Science Agenda, American Psychological Association*, 22.
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831-843.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389-433.
- La Brozzi, R. M. (2012). *The role of study abroad and inhibitory control on processing redundant cues*. Paper presented at the 14th Hispanic linguistics symposium, Somerville, MA.
- Ladefoged, P. (1967). *Three areas of experimental phonetics*. London: Oxford University Press.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.
- Lai, Y. H. (2012). Perceptual processing of Mandarin nasals by L1 and L2 Mandarin speakers. *J Psycholinguist Res*, 41(4), 237-252.
- Lamendella, J. (1977). General principles of neurofunctional organization and their manifestations in primary and non-primary language acquisition. *Language Learning*, 27, 155-196.

- Lee, J., Lyoo, I., Kim, S., Jang, H., & Lee, D. (2005). Intellect declines in healthy elderly subjects and cerebellum. *Psychiatry and Clinical Neurosciences*, 59, 45-51.
- Lee, M. (1976). Pronunciation problems are different and similar among Cantonese and Mandarin speakers. *TESL Reporter*, 9(4), 3-6.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Leon-Carrion, J., García-Orza, J., & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *International Journal of Neuroscience*, 114(10), 1291-1311.
- Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: expectations about word length guide infant statistical learning. *Cognition*, 122(2), 241-246.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431-461.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.
- Lieberman, H., Faaborg, A., Daher, W., & Espinosa, J. (2005). *How to wreck a nice beach you sing calm incense*. Paper presented at the International Conference on Intelligent User Interfaces (IUI 2005), San Diego.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13-44). Orlando, FL: Academic Press.
- Lindsay, S., & Gaskell, M. G. (2009). *Spaced learning and the lexical integration of novel words*. Paper presented at the 31st annual conference of the cognitive science society, Austin, TX.
- Lindsay, S., & Gaskell, M. G. (2010). A complementary system account of word learning in L1 and L2. *Language Learning*, 60(2), 45-63.
- Linton, L. R., & Harder, L. D. (2007). *Biology 315—Quantitative Biology Lecture Notes*. Calgary, AB: University of Calgary.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of Acoustical Society of America*, 89, 874-886.
- Long, M. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics in Language Teaching*, 43(4), 287-317.
- Long, M. H. (1990). Maturation constraints on language development. *Studies in Second Language Acquisition*, 12(3), 251-285.
- Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: a review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, 2, 174-207.
- Mackey, A., Philip, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. . In P. Robinson (Ed.), *individual differences and instructed language learning* (pp. 181-210): John Benjamins Publishing Company.
- Mackey, A., & Sachs, R. (2012). Older learners in SLA research: a first look at working memory, feedback, and L2 development. *Language Learning*, 62(3), 704-740.
- MacWhinney, B. (1987). *The Competition Model*. Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31(4), 883-914.
- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. de Groot (Eds.), *Handbook of Bilingualism* (pp. 49-67). New York: Oxford University Press.

- MacWhinney, B. (2008). A Unified Model. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, 34(1), 9-34.
- Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385, 813–815.
- Mårtensson, J., Eriksson, J., Bodammer, N.C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *Neuroimage*, 63, 240-244.
- Martini, M., Furtner, M. R., & Sachse, P. (2013). Working memory and its relation to deterministic sequence learning. *PLoS One*, 8(2), e56166.
- Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, 34, 338–348.
- McGeorge, P., Crawford, J. R., & Kelly, S. W. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 239-245.
- McLaughlin, B. (1992). Myths and misconceptions about second language learning: what every teacher needs to unlearn. *Educational Practice Report*, 5, 1-18.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat Neurosci*, 7(7), 703-704.
- McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, 60(2), 123-150.
- McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *Journal of Neuroscience*, 26(29), 7629-7639.
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39, 21-46.
- Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.
- Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Henry Holt & Co.
- Miralpeix, I. (2007). Lexical knowledge in instructed language learning: the effects of age and exposure. *International Journal of English Studies*, 7(2), 61-83.
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Language Learning*, 62(1), 302-331.
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: the role of prediction-based learning in language. *Topics in Cognitive Science*, 2(1), 138-153.
- Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers. An investigation of the preterite-imperfect contrast in Spanish. *Studies in Second Language Acquisition*, 25(3), 351–398.
- Morgan-Short, K., & Ullman, M. T. (2012). The neurocognition of second language. In S. Gass & A. Mackey (Eds.), *The routledge handbook of second language acquisition* (pp. 282-299). New York: Routledge.
- Muñoz, C. (2006). The effects of age on foreign language learning: the BAF project. . In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29 578–596.

- Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching*, 49(2), 113-133.
- Myerson, J., Ferraro, F. R., Hale, S., & Lima, S. D. (1992). General slowing in semantic priming and word recognition. *Psychology and Aging*, 7, 257-270.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull*, 12, 826-859.
- Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language*, 92, 26-32.
- Neufeld, G. (1978). On the acquisition of prosodic and articulatory features in adults language learning. *Canadian Modern Language Review*, 34, 163-174.
- Neufeld, G. G. (1977). Language learning ability in adults: a study on the acquisition of prosodic and articulatory features. *Working Papers on Bilingualism*, 12, 44-60.
- Neufeld, G. G. (1979). Towards a theory of language learning ability. *Language Learning*, 29(2), 227-241.
- Neufeld, G. G. (1988). Phonological asymmetry in second-language learning and performance. *Language Learning*, 38(4), 531-559.
- Newport, E. L. (1988). Constraints on learning and their role in language acquisition: studies of the acquisition of American Sign Language. *Language Sciences*, 10, 147-172.
- Newport, E. L. (1990). maturational constraints on language learning. *Cogn Sci*, 14, 11-28.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance. I. Statistical learning of non-adjacent dependencies. *Cogn Psychol*, 48(2), 127-162.
- Nikolov, M., & Mihaljević Djigunović, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics*, 26, 234-260.
- Nikolov, M., & Mihaljević Djigunović, J. (2011). All shades of every color: an overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics*, 31, 95-119.
- Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of Acoustical Society of America*, 115(4), 1777-1790.
- Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cogn Psychol*, 34, 191-243.
- North, B., & Jones, N. (2009). Relating language examinations to the Common European Framework of Reference for languages: learning, teaching, assessment (CEFR). *Language Policy Devison, Strassbourg*.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411-421.
- Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., & Hagiwara, H. (2011). Age and amount of exposure to a foreign language during childhood: behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neuroscience Research*, 70(2), 197-205.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13-B23.

- Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: the neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neuroscience*, *24*, 8436–8440.
- Ortega, L. (2013). SLA for the 21st century: disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, *63*, 1-24.
- Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs and event-related potentials: a means for exploring the neurocognition of second-language processing. *Language Learning*, *56*, 199–230.
- Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., & Pitkänen, I. (2008). Second language learning and changes in the brain. *Journal of Neurolinguistics*, *21* (6), 509-521.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258-278.
- Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the hebb repetition effect and the learning of phonological word-forms. *Philosophical Transactions of the Royal Society B*, *364*, 3737-3753.
- Paradis, N. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*, 299-320.
- Park, D. C., & Payer, D. (2006). Working memory across the adult lifespan. In E. Bialystock & F. Craik (Eds.), *Lifespan cognition: mechanisms of change*. New York.: Oxford University Press.
- Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., Reggers, J., . . . Maquet, P. (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron*, *44*(3), 535-545.
- Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604-607.
- Penfield, W., & Roberts, L. (1959). *Speech and brain-mechanisms*. Princeton, NJ: Princeton University Press.
- Perdue, C. (1996). Pre-basic varieties: the first stages of second language acquisition. *Toegepaste taalwetenschap in artikelen*, *55*, 135-150.
- Perquin, M., Vaillant, M., Schuller, A. M., Pastore, J., Dartigues, J. F., Lair, M. L., & Diederich, N. (2013). Lifelong exposure to multilingualism: new evidence to support cognitive reserve hypothesis. *PLoS One*, *8*(4), e62030.
- Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: extracting word-like units when only statistical information is available. *Journal of Memory and Language*, *66*(4), 807-818.
- Persad, C. C. , Abeles, N., Zacks, R. T., & Denburg, N. L. (2002). Inhibitory changes after age 60 and their relationship to measures of attention and memory. *Journals of Gerontology. Series B: Psychological Sciences and Social Sciences*, *57*, 223-232.
- Pike, K.L. (1972 [1945]). The intonation of American English. In D. Bolinger (Ed.), *Intonation* (Vol. 53-83). Harmondsworth: Penguin.
- Pinker, S. (1994). *The language instinct*. New York, NY: W. Morrow and Co.
- Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, *60*(941-951).
- Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychol Bull*, *133*(2), 227-244.
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*(1), 62-88.

- Pullum, G. K. (1996). *Learnability, hyperlearning, and the poverty of the stimulus*. Paper presented at the 22nd Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Role of Learnability in Grammatical Theory, Berkeley, CA.
- Rast, R. (2008). *Foreign language input: initial processing*. Clevedon: Multilingual Matters.
- Rast, R., & Dommergues, J.-Y. (2003). Towards a characterisation of saliency on first exposure to a second language. In S. Foster-Cohen & S. Pekarek-Doehler (Eds.), *EUROSLA Yearbook 3* (pp. 807-810). Amsterdam: Benjamins.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical decision task. *Psychology and Aging, 19*(2), 278-289.
- Raven, J. (1962). *Advances progressive matrices, set II*. London: Lewis, H. K.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855-863.
- Reber, A. S., Allen, R., & Reber, P. J. (1999). Implicit versus explicit learning. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 475-513). Cambridge, MA: MIT Press.
- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 888-896.
- Reichle, R. V. (2010). Judgments of information structure in L2 French: nativelike performance and the critical period hypothesis. *International Review of Applied Linguistics in Language Teaching, 48*(1), 53-85.
- Reichle, R. V., & Birdsong, D. (in press). Processing focus structure in L1 and L2 French: L2 proficiency effects on ERP's. *Studies in Second Language Acquisition*.
- Roberts, L., Dimroth, C., & Gullberg, M. (2010). *Investigating what word form knowledge adults and children can acquire after the first few minutes of exposure to a new language*. Manuscript in preparation.
- Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*(1), 45-99.
- Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: a replication and extension of Reber, Walkenfeld and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Amsterdam: John Benjamins.
- Robinson, P. (2005). Cognitive abilities, chunk-strength and frequency effects in implicit artificial grammar and incidental L2 learning: replications of Reber, Walkenfeld and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition, 27*, 235-268.
- Robinson, P. (2010). Implicit artificial grammar and incidental natural second language learning: how comparable are they? *Language Learning, 60*(2), 245-263.
- Robinson, R., Mackey, A., Gass, S., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247-267). New York: Routledge.
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.
- Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, 35, 606-621.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2), 101-105.
- Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39, 484-494.
- Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, 8(4), 535-543.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403-428.
- Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journals of Gerontology: Psychological Sciences*, 50, 297-306.
- Schmidt, K.-H., & Metzler, P. (1992). *Wortschatztest (WST)*. Weinheim: Beltz Test GmbH.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.
- Schwartz, B. D., & Eubank, L. (1996). What is the 'L2 initial state'? *Second Language Research*, 12(1), 1-5.
- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12(1), 40-72.
- Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, 19(3), 245-253.
- Seger, C. A. (1994). Implicit learning. *Psychol Bull*, 115(2), 163-196.
- Seger, C. A., Prabhakaran, V., Poldrack, R. A., & Gabrieli, J. D. E. (2000). Neural activity differs between explicit and implicit learning of artificial grammar strings: an fMRI study. *Psychobiology*, 28(3), 283-292.
- Seidenberg, M. S., Macdonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, 298, 553-554.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209-231.
- Service, E. (1992). Phonology, working memory, and foreign language learning. *Q J Exp Psychol (Hove)*, 45A, 21-50.
- Sharwood Smith, M. S. (1986). Comprehension versus acquisition: two ways of processing input. *Applied Linguistics*, 7(3), v.
- Shoemaker, E., & Rast, R. (2013). Extracting words from the speech stream at first exposure. *Second Language Research*, 29(2), 165-183.
- Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: a novel lexicon study. *Second Language Research*, 29(2), 185-200.
- Singleton, D. (2005). The critical period hypothesis: a coat of many colours. *International Review of Applied Linguistics in Language Teaching*, 43(4), 269-285.
- Singleton, D., & Ryan, L. (2004). *Language acquisition: the age factor* (2 ed.). Clevedon, UK: Multilingual Matters.
- Slabakova, R. (2013). Adult second language acquisition: a selective overview with a focus on the learner linguistic system. *Linguistic Approaches to Bilingualism*, 13, 48-72.

- Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: evidence from second language learning. *Child Development*, 49(4), 1114-1128.
- Soper, D. S. (2013). p-value calculator for correlation coefficients. Retrieved 25.11.2013, 2013, from <http://www.danielsoper.com/statcalc>
- Sparks, R., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*, 21(90), 90-111.
- Spearman, C. (1904). 'General Intelligence', objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.
- Specht, K., Holtel, C., Zahn, R., Herzog, H., Krause, B. J., Mottaghy, F. M., . . . Huber, W. (2003). Lexical decision of nonwords and pseudowords in humans: a positron emission tomography study. *Neurosci Lett*, 345(3), 177-181.
- Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: a longitudinal electrophysiological study on second language learning. *Neuroimage*, 33(774-783).
- Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Language Research*, 25(1), 13-41.
- Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 8, 448-460.
- Student, B. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25.
- Szmalc, A., Brysbaert, M., & Duyck, W. (2012). Working memory and (second) language processing. *Memory, language, and bilingualism: Theoretical and applied approaches*, 74-94.
- Tamminen, J., & Gaskell, M. G. (2006). *Learning new words: effects of lexical competition and age of acquisition*. Paper presented at the Twenty-Eighth Annual Conference of the Cognitive Science Society, Mahwah, NJ.
- Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience*, 30(43), 14356-14360.
- Tewes, U. (1991). *HAWIE-R: Hamburg Wechsler Test für Erwachsene, Revision*. Bern: Huber.
- Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. *Journal of Cognitive Neuroscience*, 16(8), 1339-1351.
- Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control: when a little frontal lobe goes a long way. *Association for Psychological Science*, 18(5), 259-263.
- Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: an event-related potential investigation. *Studies in Second Language Acquisition*, 27, 173-204.
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, B25-B34.
- Tremblay, K. L., Shahin, A., Picton, T., & Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clinical Neurophysiology*, 120, 128-135.
- Truscott, J. (1998). Noticing in second language acquisition: a critical review. *Second Language Research*, 14, 103-135.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.

- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231-270.
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: examining the correlation between operation span and raven. *Intelligence*, 33, 67-81.
- van Boxtel, S., Bongaerts, T., & Coppen, P.-A. (2005). Native-like attainment of dummy subjects in Dutch and the role of the L1. *International Review of Applied Linguistics in Language Teaching*, 43(4), 355–380.
- van Hell, J. G., & Tanner, D. (2012). Second Language Proficiency and Cross-Language Lexical Activation. *Language Learning*, 62(s2), 148-171.
- Vanhove, J. (2014). *Receptive multilingualism across the lifespan: cognitive and linguistic factors in cognate guessing*. (PhD), University of Fribourg, Fribourg (Switzerland).
- VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.
- Veroude, K., Norris, D. G., Shumskaya, E., Gullberg, M., & Indefrey, P. (2010). Functional connectivity between brain regions involved in learning words of a new language. *Brain & Language*, 113(1), 21-27.
- Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech*, 40, 47-62.
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron*, 35(1), 205-211.
- Wang, Y., Spence, M.M., Jongman, A. , & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106, 3649–3658.
- Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, 4, 123–130.
- Weber, A., & Broersma, M. (2012). Spoken word recognition in second language acquisition. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Bognor Regis: Wiley-Blackwell.
- Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, 119(1), 597.
- Weber-Fox, C. M., & Neville, H. J. (1996). Maturation constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8(3), 231-256.
- White, L. (1991). Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom. *Second Language Research*, 7(2), 133-161.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*.
- Wilhelm, I., Diekelmann, S., & Born, J. (2008). Sleep in children improves memory performance on declarative but not procedural tasks. *Learning & Memory*, 15(5), 373-377.
- Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, 27, 269-304.
- Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319-353). Bingley, UK: Emerald Group Publishing.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., . . . Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, 638-642.
- Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B. , Poliva, O., . . . Reed, B. (2010). Improving digit span assessment of short-term

- verbal memory. *Journal of Clinical and Experimental Neuropsychology*, 4, 1-11.
- Yelland, G., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics*, 14(4), 423-444.
- Zhang, Y., Kuhl, P., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *Neuroimage*, 26, 703-720.
- Zwitserslood, P., Klein, W., Liang, J., Perdue, C., Kellerman, E., & Wenk, B. (1994). *The first few minutes of foreign-language exposure*. Unpublished manuscript. MPI for Psycholinguistics, Nijmegen.
- Zyzik, E. (2009). The role of the input revisited: nativist versus usage-based models. *L2 Journal*, 1, 42-61.

# Appendices

## A - The Mandarin Weather Report

*Text associated with weather chart 1*

1. rang4 wo3-men2 xian1 kan4 zhe4 ge di4-qu1  
let us first look this CLF area  
Let's first look at this area.
2. Nanhai qing2-kuang4bu4 hui4 bian4-hua4  
Nanhai situation not will change  
The situation in Nanhai won't change.
3. Lanzhou ming2-mei4 yang2-guang1 bao3-chi2 zhi4 xing1-qi1-wu3  
Lanzhou bright sunshine stay until Friday  
The bright sunshine will stay until Friday in Lanzhou.
4. gu1ji4 zhou1-mo4 hui4 gua1-feng1  
estimate weekend will blow-wind  
We predict that it will be windy at the weekend.
5. nuan3 kong1-qi4 qian2-feng1 sui2-hou4 nuo2 chu1 Lan2zhou1  
warm air front afterwards move exit Lan2zhou1  
The warm front will then move out of Lan2zhou1
6. ta1 man4-man4 yi2-dong4  
it slowly move  
moving slowly
7. cong2 nan2-fang1 dai4-lai2 shi1-run4 qi4-liu2  
fromsouth bring-come moist air-current  
bringing moist currents from the south
8. Lanzhou yin1-ci3 xia4-yu3  
Lanzhou therefore rain  
It will therefore rain at Lanzhou.
9. san1ri4 hou4 yi2-dong4 xun4-su4  
three days later move quickly  
Three days later it will move quickly
10. yun2 kuo4-san4 dao4 fu4-jin4 cheng2-shi4 Yin2Chuan1  
cloud spreadto neighbor city Yin2Chuan1  
and clouds will spread to the neighbor city Yinchuan.
11. zhe4pian4 di4-qu1 zhi1-qian2 bei4 leng3 kong1-qi4 zhan4-ju4  
this CLF area beforePASS cold air occupy  
A cold current occupies this area since before
12. liang3 zhong3 qi4-liu2 ji2-zhong1  
two kind air-current gather  
Two kinds of currents gather here
13. zhao4-cheng2 lei2 dian4 jiao1-jia1  
cause thunder lightning mix  
and this causes thunder and lightning.
14. mei3 ge nan2-fang1 cheng1-shi4 dou1 shou4 ying3-xiang3  
every CLF southern cities all receive  
influence  
Every southern city will be affected.
15. xiao3 hai2-zi-men zhu4-yi4 bu4 dai1 zai4 hu4-wai4  
little children pay-attention not stay at outdoors  
Small children should not stay outdoors.
16. dan4 wo3-men2 reng2-jiu4 ke3-yi3 gan3-shou4 pian4-ke can4-lan4 yang2-  
guang1

but we still can enjoy moment brilliant sun light  
 But we can still enjoy a moment of sunshine.

*Text associated with weather chart 2*

17. xian4-zai4 zhuān3-xiāng4      duì4-miàn4      Lǐ3zhuāng1      zhè4      piān4.  
 now      turn      opposite      Lizhuang      this  
 area  
 now let's turn to this area Lizhuang.
18. shān1 shāng4      qí4-wēn1      tu1-ran2      zhuān3      dī1  
 mountain top      temperature suddenly      turn      low  
 The temperature will drop at the mountain top.
19. shēn4-zhī4      kě3-néng2      dī1      guó4      líng2      dù4  
 even      maybe      low      than 0      degree  
 It may even be below zero.
20. rén2-mén2      míng2-xiān3      gān3-dào4      hēn3      hán2-léng3  
 people      obviously      feel      very      cold  
 People will obviously feel very cold
21. léng3      kōng1-qí4      qián2-fēng1      zhān4-jū4      Lǐ3zhuāng1  
 cold air      front      occupy      Lizhuang  
 The front of the cold current occupies Lizhuang,
22. dài4-lái2      bú4      tài4      hào3      xíng2-shì4  
 bring come not      too      good situation  
 bringing a not-too-good situation.
23. Lǐ3zhuāng1      yīn1-cǐ3      jiāng4-xuě3  
 Lizhuang      therefore      snow  
 It will therefore snow at Lizhuang
24. jī2-zhōng1      yú2      běi3      qū1  
 center      at      north area  
 mainly in the north.
25. qián2-fēng1      zài4      Lǐ3zhuāng1      tíng2-liú2      yī1      tiān1  
 front      at      Lizhuang      stop-stay one day  
 The front will stay over Lizhuang for one day.
26. tóng2-shí2      huì4      guā1-fēng1,  
 same time      will blow-wind  
 It will be windy at the same time.
27. gěi3      měi3      gě      dì4-fāng1      dài4-lái2      jiāo1-tōng1      kūn4-nán2  
 give all      CLF      places bring      traffic      trouble  
 It causes traffic troubles in all districts.
28. zhè4zhōng3      qīng2-xíng2      chí2-xú4      jiāo3      cháng2      shì2-jian1  
 this      kind      situation      last      rather long      time  
 This kind of situation will last for a rather long time
29. shēn4-zhī4      jì4-xu4zhī4      xià4      zhōu1      yī1  
 even      extend      to      next Monday  
 and will even last till next Monday.
30. zhī1-hòu4      cān4-lán4      yáng2-guāng1      jiāng1      chū1-xiān4  
 that-after      brilliant      sun-shine      will      shine  
 After that brilliant sunshine will shine.
31. wēn1-dù4      yě3      kēn3-dīng4      zhuān3      hào3  
 temperature      also      surely      change good  
 The temperature will surely also improve
32. bǐ3      zuó2-tiān1      shēng1-gāo1      sì4      dù4  
 comparing yesterday      rise-high      4 degrees  
 In comparison to yesterday it will rise by 4 degrees

*Text associated with weather chart 3*

33. rāng4      wǒ3-mén2      kān4-kān4      yān2-hǎi3      yī1-dài4  
 let      us      see-see      coast part  
 Let's have a look at the coastal area.
34. qīng2-lǎng3      tiān1-qí4      cōng2      běi3      kuò4-sān4

- fine weather from south spread  
Fine weather will spread from the south
35. qi4-wen1 pu3-bian4 sheng1-gao1 san1 du  
temperature all rise three degrees  
The temperature will rise by three degrees.
36. xing1-qi1-er4 Zhoucun you3 qing1 wu4  
Tuesday Zhoucun have light fog  
There will be a light fog on Tuesday at Zhoucun.
37. shi1qi4 ji2-zhong1yu2 Zhoucun nan fang1  
wet air gather at Zhoucun south direction  
The moist air will gather south of Zhoucun.
38. tian1-qi4 hen3 nuan3-he2  
weather very warm  
The weather will be very warm
39. shan1 shang4 qi4-wen1 shao1-wei1 di1 liang2 du4  
mountain top temperature slightly low two  
degree  
It will be two degrees lower at the mountain top.
40. yang2-guang1 di4 er4 tian1 gan3-zou3 shi1-qi4  
sun-shine the second day dispel moisture  
The sunshine the next day will dispel the moisture
41. hu2 bing1 jiang1 man4-man4 rong2-hua4  
lake ice will slowly melt  
The ice on the lake will slowly melt
42. xing1-qi1-si4 tian1-kong1 jiang1 long3-zhao4 zhe yun2  
Thursday sky cover ASP cloud  
The sky will be covered by clouds on Thursday.
43. zhu3-yao3 ji2-zhong1 yu2 dong1 fang1.  
major gather at east direction  
The clouds will gather in the east.
44. huai4 tian1-qi4 Hou4 liang3 ri4 cong2 dong4 kai4-shi3  
bad weather after two days from east begin  
Bad weather will form in the east
45. hai2 ke3-neng2 xia4-yu3  
still probably rain  
It will probably rain
46. xi1-wang4 peng2-you2-men2 bu4 yao4 wang4-ji4 dai4 san3  
hope friends not will forget bring  
umbrella  
We hope you will not forget to bring umbrellas.
47. xian4-zai4 lai2 kan4 fu4-jin4 ban4-dao3 qing2-xing2  
now come look neighboring peninsula situation  
Let's now look at the situation in the neighboring peninsula.
48. Chang2sha1 ye3 you3 tong2-yang4 qing2-xing2  
chang2sha1 also have same situation  
The situation will be the same at Changsha.
49. qing1 wu4 gei3 ren2 qi2-miao4 gan3-jue2  
light fog give people good feeling  
A light fog will make people feel good
50. ta1 ji2-zhong1 zai4 Chang2sha1 zhong1-xin1  
it gather at Chang2sha1  
it gathers at the center of Changsha
51. sui2-hou4 shi4 can4-lan4 yang2-guang1  
afterwards be brilliant sunshine  
Afterwards there will be brilliant sunshine.
52. ran2-hou4 xia4-yu3 yi1 tian1  
afterwards rain one day

- It will then rain for one day
53. jia1-za2 zhe lei2  
mix ASP thunder  
Thunder will mix with the rain.
54. dan4 qi4-wen1 bu4 hui4 bian4-hua4 tai4 dao4  
but temperature not will change too much  
But the temperature will not change too much
55. wo3-men2 yu4-ce4 ban4-dao3 you3 duan3 shi2 bing1bao2  
we predict peninsula have short time hail  
We predict a short period of hail in the peninsula.
56. yan2-hai3 qi2-ta1 cheng2-shi4 qi4-hou4 dou1 bi3-jiao3 yi2-ren2  
coastal other cities weather all rather  
pleasant  
Other coastal cities will all have pleasant weather

*Text associated with weather chart 4*

57. da4-jia1 lai2 guan1-zhu4 zhe4 pian4 di4-qu1  
we come pay attention this piece area  
Let's turn to this area
58. yi1 gu3 han2-liu2 cong2 Mongo ru4-qin1 Meizhou  
a gust cold-current from Mongo invade Meizhou  
A cold current from Mongo will invade Meizhou.
59. liang3 ri4 nei4 xun4-su4 yi2-dong4  
two days within fast move  
It moves fast in two days
60. xiang4 dong1 zhu2-jian4 fa1-zhan3.  
towards East gradually develop  
gradually moving towards the east.
61. qian2-feng1 ming2-tian1 wan3-shang4 dao4-da2 Meizhou  
cold front tomorrow night reach Meizhou  
The cold front will reach the east of Meizhou tomorrow night
62. ta1 yi1-zhi2 xiang4 dong1 liu2-dong4  
it continuously towards east float  
It will move continuously toward the east
63. xing1-gi1-er4 hui4 jiang4-xue3  
Tuesday will snow  
It will snow on Tuesday.
64. yun2 zhu2-jian4 dui1-ji2  
cloud gradually accumulate  
clouds will gradually accumulate
65. zhu3-yao4 ji2-zhong1 yu2 Meizhou dong1 fang1  
mainly center at Meizhou east  
The clouds will center mainly to the east of Meizhou.
66. shan1 li3 bian4 de te4-bie2 han2-leng3  
mountain inside change de extremely cold  
It will be extremely cold in the mountains
67. tong2-shi2 hai2 gua1-feng1  
same time still blow-wind  
and at the same time it will be windy.
68. sui2-shi2 ke3-neng2 fa1-sheng1 xia4-yu3  
any time probably rain  
It will probably rain at any time
69. han2-liu2 huan3-man4 yi2-dong4  
cold-air slowly move  
with the cold air moving slowly
70. zhe4zhong3 qing2-kuang4yao4 chi2-xu4 ji3 tian1  
this kind situation will last several  
days



- The front will move slowly
89. cong2 cheng2 xi1 kai1-shi3  
From city west begin  
from the west of the city
90. xiang4 Dongguan fa1-zhan3  
towards Dongguan develop  
moving towards Dongguan.
91. da4 pian4 yun2 jin3-jin3 gen1-sui2  
great hosts of clouds closelyfollow  
closely followed by clouds
92. ta1-men2 huan3-man4 yi2-dong4  
they slowly move  
moving slowly
93. Dongguan yin1-ci3 xia4-yu3  
Dongguan therefore rain  
and causing rain in Dongguan.
94. ke3-neng2 yao4 gua1-feng1  
probably will blow-wind  
It will probably be windy.
95. wen1-du4 bu4 hui4 bian4.  
temperature not will change  
The temperature will not change
96. ke3 neng2 yi-tian1 duo1 gua1-feng1  
Xxx one-day blow-wind  
The wind will last 24 hours.
97. ming2-tian1 bang4-wan3 hai2 you3 duan3-zhan3 yang2-guang1  
today dusk still have short sun-  
shine  
There will still be a short period of sunshine at dusk tomorrow.
98. zhan4-shi2 gan3-zou3 yin1-chen2-chen de yun2  
temporarilydrive-away gloomy de clouds  
temporarily driving away the gloomy clouds.

*Text associated with weather chart 6*

99. rang4 wo3-men2 xiang4 xia4 kan4  
let us towards down look  
Let's look further down.
100. yin1 leng3 qi4-liu2 yi2-dong4  
for cold current move  
Because of the movement of the cold air
101. Dongguan tong-yang xia4-yu3  
Dongguan similarly rain  
it will also rain in Dongguan
102. er3-qie3 hai2 gua1-feng1  
and will blow-wind  
and it will be windy.
103. te4-bie2 shi4 shan1 shang4  
especially be mountain area  
especially in the mountain area
104. qiang2 feng1 jiang1 cong2 bei3 yi2-zhi2 chui1  
strong wind will from north continuously blow  
A strong wind will blow continuously from the north.
105. ci3 di4 de peng2-you3 zui4 hao3 bu4 yao4 chu1 men2  
this area de friends best not exit door  
people living in this area had better not go outside
106. deng3 qian2-feng1 nuo2 zou3  
wait front move away  
but wait for the front to move away.

- 107.shan1 li3 qing2-kuang4bi4-ding4 hao3 zhuan3  
 mountain inside situation must good turn  
 The situation in the mountain area will surely improve.
- 108.he2 bing1 rong2-hua4  
 river ice melt  
 The ice in the river will melt.
- 109.ting2-zhi3 lian2-xu4 xia4-yu3  
 stop continuous raining  
 and the continuous raining will stop.
- 110.ke3-yi3 fa1-xian4 yi1-dian3 fu2 yun2  
 can find a bit floating cloud  
 We may still find some floating clouds.
- 111.wei1-feng1 cong2 nan2 chui1 lai2  
 breeze from south blow-come  
 A breeze will come from the south
- 112.xiang4 bei3 chui1 qu4  
 towards north blow go  
 blowing towards the north
- 113.shan1 li3 ju1-min2 neng2 gan3-shou4 wen1-nuan3  
 mountain inside people can enjoy warmth  
 People in the mountains will enjoy the warmth.
- 114.sui2-shi2 dou1 gua1-feng1  
 anytime all blow-wind  
 The wind will blow from time to time
- 115.nuan3-liu2 kuan4-su4 yi2-dong4  
 warm-current fast move  
 The warm current will move fast
- 116.dan4 wen1-du4 bu4 hui4 fei1-chang2 di1  
 but temperature not will very low  
 but the temperature will not be very low.
- 117.zui4-hou4 wo3-men2 xiang4 xi1 guan1-zhu4 Hanyang qing2-kuang4  
 lastly us towards west attend Hanyang  
 situation  
 Lastly let's look at the situation in the west in Hanyang
- 118.ci3 di4-qu1 guo4-qu4 ji3 zhou1 yi1-zhi2 gua1-feng1  
 this area past several week alwaysblow-wind  
 It has been constantly windy in this area in the past several weeks.
- 119.jin1 ming2 you3 nuan3-liu2 yi2-dong4 dao4 ci3 di4.  
 today tomorrow have warm-current move to  
 this  
 A warm current will move to this area today and tomorrow
- 120.ren2-men2 ke3-yi3 xiang3-shou4 can4-lan4 yang2-guang1  
 people can enjoy brilliant sunshine  
 so people can enjoy some brilliant sunshine.

## B – The Chinese Lexical Decision Task

256 monosyllables from Chinese Lexical Decision Task (LDT) with tone marks

### *Non-words (analyzed)*

VCCC	VCC	CV_nasal	CV_plosive
airsp2	ains4	maim4	mait1
aorsp2	ans3	mam3	maot1
arsp1	aons4	maom4	mat4
ersp1	ens3	mem3	met4
oursp3	ouns1	moum1	mout2
uirsp4	uins2	muim2	muit3
uorsp4	uns1	mum1	muot3
ursp3	uons2	muom2	mut2
ailst3	ailp1	gaim4	gaip1
alst2	alp4	gam3	gaop1
aolst3	aolp1	gaom4	gap4
elst2	elp4	gem3	gep4
oulst4	oulp2	goum1	goup2
uilst1	uilp3	guim2	guip3
ulst4	ulp2	gum1	guop3
uolst1	uolp3	guom2	gup2

### *Non-words (not analyzed)*

CCCV	CCV	Pseudo-words	
schra1	sna3	biong3	buai1
schrai2	snai4	do1	fai3
schrao2	snao4	ko3	mong4
schre1	sne3	lueng1	piang3
schrou3	snou1	niong2	ra1
schru3	snu1	no2	suang4
schruai4	snui2	tueng4	ten2
schruo4	snuo2	zhiao4	xa2
spra2	bla4	be1	chueng3
sprai3	blai1	cei3	fong4
sprao3	blao1	chei4	gi1
spre2	ble4	len3	hiu3
sprou4	blou2	pe2	kia2
spru4	blu2	shong1	muo1
sprui1	blui3	tiu2	rei2
spruo1	bluo3	zuai4	ruang4

### *Real words (fillers that were heard for the first time in the LDT)*

shuang3	zhuai4	pian2	sao3
se4	beng3	mi2	hun1
kui2	gun3	tang2	ma3
keng1	shu3	lia3	kua1
cao2	geng4	lie4	ba2
lin2	huang4	gang1	cuo1
niu2	mang3	jing4	xiong1
rui4	song4	gou3	teng2
juan4	ben1	pen1	she2
yong4	zang4	jie3	chuai4
chun3	kuai4	suo3	ni3
zan1	ruan3	zhuo2	dian1
qun1	diao1	niang4	lou1
pei2	fan2	sai4	dang4
pai3	pin1	ang2	qiu1

na2	heng2	bo1	sou3
<i>Real words (fillers that have appeared in WR)</i>			
In WR one time	In WR two times	In WR three times	In WR four times
huo2	ji3	san4	jiu4
kong1	nuo2	gei3	huan3
pu3	gu3	dou1	su4
yao3	guan1	peng2	lan4
duo1	wei1	he2	xing2
che4	ting2	chui1	dong1
ling2	wu4	zhong3	han2
ken3	guo4	shang4	zhuang1
gen1	gao1	zui4	zhu2
fang2	bie2	xin1	xing1
miao4	duan3	bing1	nan2
sa3	zhao4	shi3	bei3
run4	lian2	ran2	fa1
bao2	tai4	tong2	zou3
bang4	qin1	ri4	can4
gai1	wan3	liang3	jiao3

## C – Statistical Analysis: additional Figures and Tables

### Data Analysis Study 1 - Cross-validation

#### Accuracy

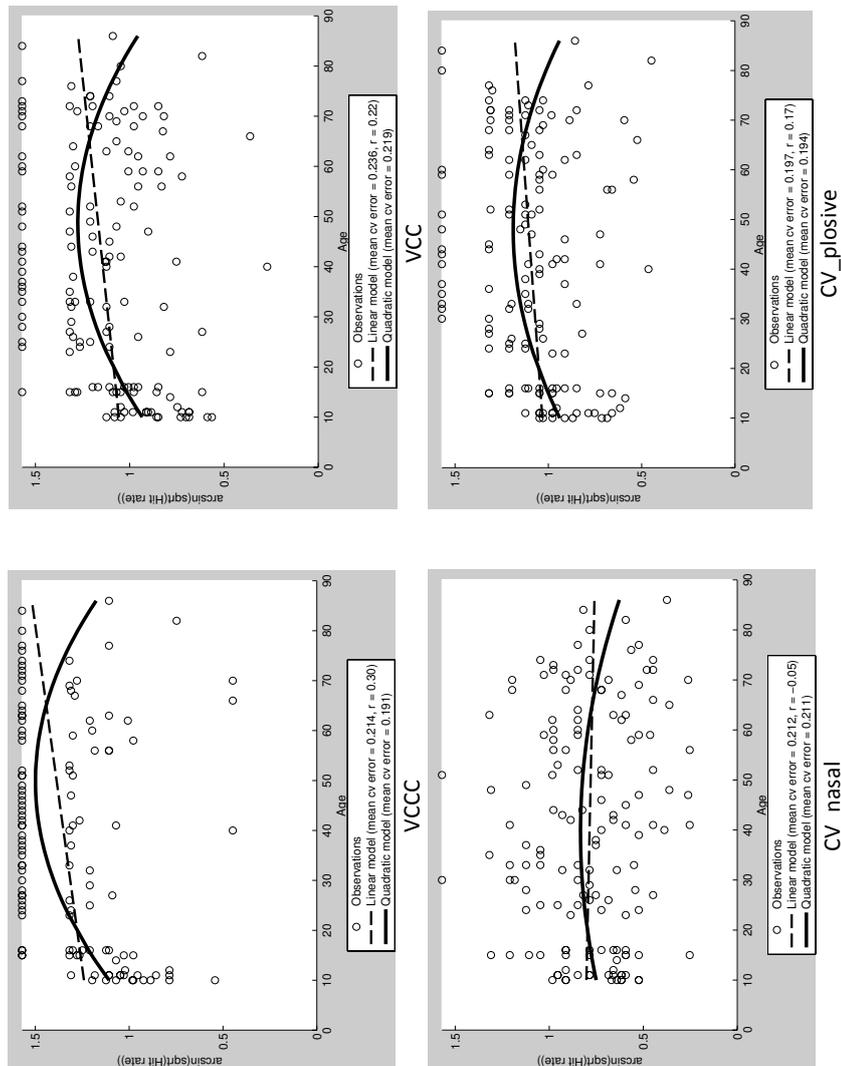


Figure 30: Cross-validation plots for all six syllable conditions, comparing a linear and quadratic model, in terms of the arcsine of the mean hit rate, across the lifespan.

**Reaction times**

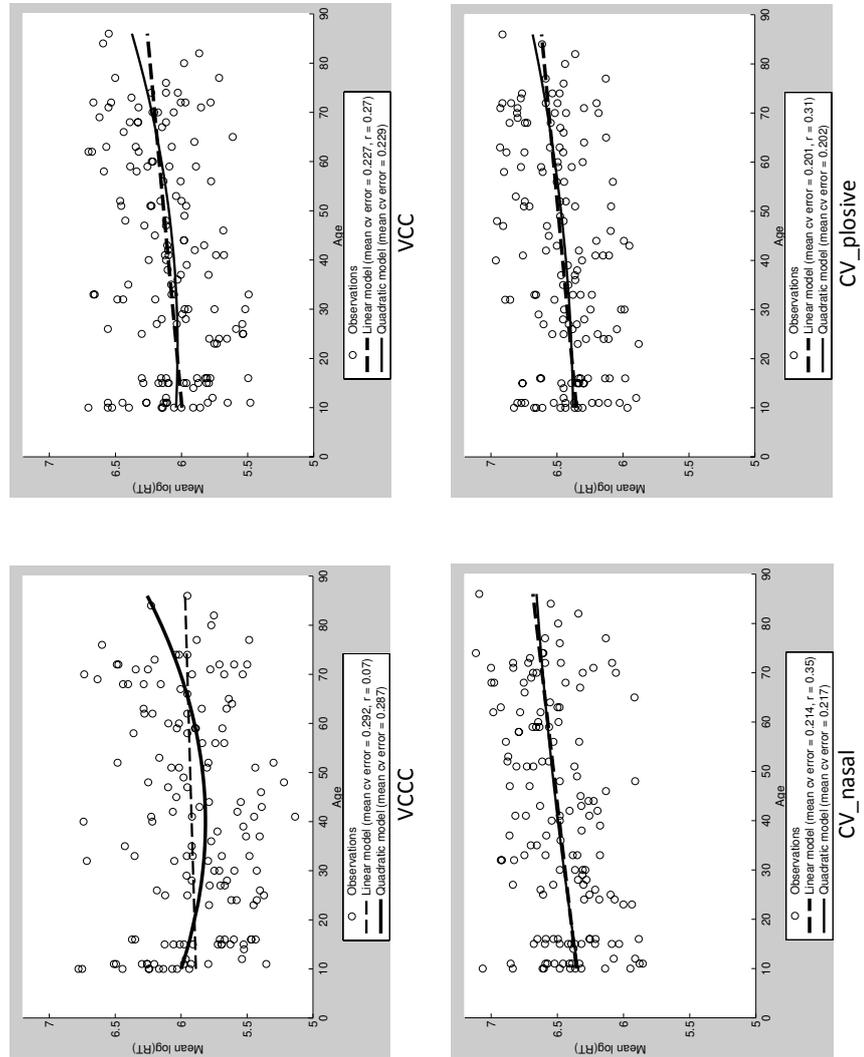


Figure 31: Cross-validation plots for all six syllable conditions, comparing a linear and quadratic model, in terms of the log of mean reaction times to correct answers, across the lifespan.

## ANOVA Study 1

### *Accuracy*

Table 18: Analysis of Variance to compare the mean transformed hit rates of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive).

	df	Sum Sq	Mean Sq	F	Prob>F	
Syllable Cond.	3	25.0	8.33	111	0.0000	***
Error	604	45.5	0.08			
Total	607	70.5				Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

### *Reaction times*

Table 19: Analysis of Variance to compare the mean transformed reaction times of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive).

Source	df	Sum Sq	Mean Sq	F	Prob>F	
Syllable Cond.	3	34.9	11.6	131	0.0000	***
Error	604	53.8	0.09			
Total	607	88.8				Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

## ANOVA Study 2

### *Accuracy – words-non-words; between groups, time 1*

Table 20: Omnibus ANOVA to compare the mean transformed hit rates of the different groups for words and non-words (VCCC, VCC, CV\_nasal, CV\_plosive).

	df	Sum Sq	Mean Sq	F	Prob>F	
Group	1	0.01	0.01	0.53	0.4681	
Age	1	0.53	0.53	33.5	0.0000	***
WordType	1	0.03	0.03	1.77	0.1858	
Group * Age	1	0.00	0.00	0.01	0.9132	
Group * WordType	1	0.00	0.00	0.21	0.6459	
Age * WordType	1	0.22	0.22	13.9	0.0003	***
Error	129	2.02	0.02			
Total	135	2.84				Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

### *Reaction times – words-non-words; between groups, time 1*

Table 21: Omnibus ANOVA to compare the mean transformed reaction times of the groups for words and non-words (VCCC, VCC, CV\_nasal, CV\_plosive).

	df	Sum Sq	Mean Sq	F	Prob>F	
Group	1	0.49	0.49	11.9	0.0008	***
Age	1	0.00	0.00	0.01	0.9247	
WordType	1	0.36	0.36	8.85	0.0035	**
Group * Age	1	0.00	0.00	0.11	0.7428	
Group * WordType	1	0.00	0.00	0.03	0.8613	
Age * WordType	1	0.02	0.02	0.48	0.4888	
Error	129	5.30	0.04			
Total	135	6.19				Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

***Accuracy – non-words; across groups, time 1***

Table 22: ANOVA to compare the mean transformed hit rates of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive) across the four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1).

	df	Sum Sq	Mean Sq	F	Prob>F	
Syllable Cond.	3	8.11	2.70	44.3	0.0000	***
Error	268	16.4	0.06			
Total	271	24.5				

Signif. codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05

***Reaction times – non-words; across groups, time 1***

Table 23: ANOVA to compare the mean transformed reaction times of the different syllable conditions (VCCC, VCC, CV\_nasal and CV\_plosive) across all four groups (Control Children, Control Adults, Experimental Children time 1, Experimental Adults time 1).

	df	Sum Sq	Mean Sq	F	Prob>F	
Syllable Cond.	3	6.78	2.26	34.6	0.0000	***
Error	268	17.5	0.07			
Total	271	24.3				

Signif. codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05

***Accuracy – non-words; between groups, time 1***

Table 24: Omnibus ANOVA to compare the mean transformed hit rates of the different groups per syllable conditions (VCCC, VCC, CV\_nasal, CV\_plosive).

	df	Sum Sq	Mean Sq	F	Prob>F	
Group	1	0.06	0.06	1.57	0.2111	
Age	1	6.86	6.86	166	0.0000	***
SyllableCondition	3	10.6	3.53	85.3	0.0000	***
Group * Age	1	0.49	0.49	11.9	0.0006	***
Group * SyllableCondition	3	0.24	0.08	1.97	0.1183	
Age * SyllableCondition	3	1.61	0.54	13.0	0.0000	***
Error	463	19.2	0.04			
Total	475	39.2				

Signif. codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05

***Reaction times – non-words; between groups, time 1***

Table 25: Omnibus ANOVA to compare the mean transformed reaction times of the groups per Syllable condition (VCCC, VCC, CV\_nasal, CV\_plosive).

	df	Sum Sq	Mean Sq	F	Prob>F	
Group	1	1.73	1.73	29.1	0.0000	***
Age	1	0.05	0.05	0.80	0.3718	
SyllableCondition	3	9.73	3.24	54.6	0.0000	***
Group * Age	1	0.01	0.01	0.19	0.6660	
Group * SyllableCondition	3	0.09	0.03	0.51	0.6785	
Age * SyllableCondition	3	1.29	0.43	7.23	0.0001	***
Error	463	27.5	0.06			
Total	475	40.7				

Signif. codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05

**Accuracy times – non-words; time 1 versus time 2**

Table 26: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given VCCC.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.001	0.001	0.017	0.8968	
Age	1	2.76	2.76	37.5	0.0000	***
Time * Age	1	0.33	0.33	8.47	0.0067	**
Error	30	1.17	0.039			
Total	63	6.47	Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05			

Table 27: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given VCC.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.14	0.14	2.60	0.1171	
Age	1	1.66	1.66	22.6	0.0000	***
Time * Age	1	0.057	0.057	1.10	0.3024	
Error	30	1.55	0.052			
Total	63	5.61	Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05			

Table 28: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given CV\_nasal.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.012	0.012	0.41	0.5286	
Age	1	1.085	1.09	12.7	0.0012	**
Time * Age	1	0.11	0.11	3.69	0.0643	
Error	30	0.89	0.030			
Total	63	4.65	Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05			

Table 29: Repeated-measures ANOVA for mean transformed hit rate over experimental groups (Children versus Adults), time 1 versus time 2, given CV\_plosive.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.001	0.001	0.001	0.9953	
Age	1	3.92	3.92	60.6	0.0000	***
Time * Age	1	0.001	0.001	0.034	0.8560	
Error	30	0.83	0.028			
Total	63	6.70	Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05			

**Reaction times – non-words; time 1 versus time 2**

Table 30: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given VCCC.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.12	0.12	6.69	0.0148	*
Age	1	0.85	0.85	20.4	0.0001	***
Time * Age	1	0.001	0.001	0.021	0.8855	
Error	30	0.55	0.018			
Total	63	2.77	Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05			

Table 31: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given VCC.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.25	0.25	8.80	0.0059	**
Age	1	1.11	0.11	1.93	0.1746	
Time * Age	1	0.001	0.001	0.022	0.8831	
Error	30	0.85	0.028			
Total	63	2.83				Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

Table 32: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given CV\_nasal.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.025	0.025	0.70	0.4106	
Age	1	1.54	0.54	5.80	0.0224	*
Time * Age	1	0.036	0.036	1.03	0.3195	
Error	30	1.06	0.035			
Total	63	4.48				Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

Table 33: Repeated-measures ANOVA for mean transformed reaction times over experimental groups (Children versus Adults), time 1 versus time 2, given CV\_plosive.

	df	Sum Sq	Mean Sq	F	Prob>F	
Time	1	0.046	0.046	1.16	0.2907	
Age	1	0.12	0.11	1.88	0.1803	
Time * Age	1	0.001	0.001	0.005	0.8560	
Error	30	1.19	0.040			
Total	63	3.20				Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05