

# Receptive multilingualism across the lifespan

*Cognitive and linguistic factors  
in cognate guessing*

Jan Vanhove

Dissertation zur Erlangung der Doktorwürde  
an der Philosophischen Fakultät  
der Universität Freiburg in der Schweiz

Genehmigt von der philosophischen Fakultät auf Antrag der Professoren  
Raphael Berthele (1. Gutachter) und Charlotte Gooskens (2. Gutach-  
terin).

Freiburg, den 17. Januar 2014.

Prof. Marc-Henry Soulet, Dekan.

2014



# Receptive multilingualism across the lifespan

*Cognitive and linguistic factors  
in cognate guessing*

Jan Vanhove

Cite as:

Vanhove, Jan (2014). *Receptive multilingualism across the lifespan. Cognitive and linguistic factors in cognate guessing*. PhD thesis. University of Fribourg (Switzerland).

Data and computer code available from:

<http://dx.doi.org/10.6084/m9.figshare.795286>.



# Contents

<b>Tables</b>	<b>xi</b>
<b>Figures</b>	<b>xiii</b>
<b>Preface</b>	<b>xv</b>
<b>I Introduction</b>	<b>1</b>
<b>1 Context and aims</b>	<b>3</b>
1.1 Cross-linguistic similarities in language learning . . . . .	4
1.2 Receptive multilingualism . . . . .	5
1.3 Multilingualism and the age factor . . . . .	8
1.4 The present project . . . . .	9
1.4.1 The overarching project ‘Multilingualism through the lifespan’ . . . . .	9
1.4.2 Aim, scope and terminology . . . . .	10
1.5 Overview . . . . .	13
<b>II The lifespan development of cognate guessing skills</b>	<b>15</b>
<b>2 Inter-individual differences in cognate guessing skills</b>	<b>17</b>
2.1 Linguistic repertoire . . . . .	18
2.1.1 Typological relation between the L $x$ and the L1 . . . . .	18
2.1.2 The impact of multilingualism . . . . .	19
2.2 Previous exposure . . . . .	26

2.3	Attitudes . . . . .	28
2.4	Age . . . . .	29
<b>3</b>	<b>The lifespan development of cognition</b>	<b>33</b>
3.1	Intelligence . . . . .	34
3.1.1	Fluid and crystallised intelligence . . . . .	34
3.1.2	Lifespan trajectories . . . . .	36
3.1.3	Implications . . . . .	38
3.2	Working memory . . . . .	39
3.2.1	The Baddeley–Hitch multi-component model . . . . .	40
3.2.2	Measuring working memory capacity . . . . .	43
3.2.3	Lifespan trajectories . . . . .	44
3.2.4	Implications . . . . .	45
3.3	Cognitive control . . . . .	46
<b>4</b>	<b>Method</b>	<b>49</b>
4.1	Participants . . . . .	49
4.2	Tasks and procedure . . . . .	50
4.2.1	Language background questionnaire . . . . .	51
4.2.2	German vocabulary test . . . . .	52
4.2.3	English language proficiency test . . . . .	53
4.2.4	Backward digit span task . . . . .	53
4.2.5	Raven’s advanced progressive matrices . . . . .	54
4.2.6	Cognate guessing task . . . . .	54
4.2.7	Measures not used in the analyses . . . . .	58
4.3	Method of analysis: Mixed-effects modelling . . . . .	59
4.3.1	A gentle introduction to the generalised linear mixed model . . . . .	59
4.3.2	Generalised additive models . . . . .	63
<b>5</b>	<b>Data inspection</b>	<b>67</b>
5.1	Cognate guessing data . . . . .	67
5.2	Inspection of the linguistic and cognitive variables . . . . .	70
5.2.1	Self-assessed language skills and number of foreign languages . . . . .	70
5.2.2	English language proficiency test . . . . .	71
5.2.3	German vocabulary test . . . . .	73
5.2.4	Raven’s advanced progressive matrices . . . . .	73
5.2.5	Backward digit span task . . . . .	73

5.3	Missing data imputation . . . . .	74
5.4	Multicollinearity assessment . . . . .	75
<b>6</b>	<b>Age trends in cognate guessing success</b>	<b>77</b>
<b>7</b>	<b>The impact of language skills and cognitive characteristics</b>	<b>85</b>
7.1	Written items . . . . .	85
7.2	Spoken items . . . . .	90
7.3	Variable-by-modality interactions . . . . .	91
<b>8</b>	<b>Discussion</b>	<b>95</b>
8.1	Age trends . . . . .	95
8.2	Linguistic and cognitive predictors . . . . .	96
8.3	Residual age trends . . . . .	100
8.4	Outlook . . . . .	101
<b>III</b>	<b>Integrating item-related characteristics</b>	<b>103</b>
<b>9</b>	<b>Item-related determinants of cognate guessing</b>	<b>105</b>
9.1	Previous findings . . . . .	106
9.1.1	Formal distance between cognates . . . . .	106
9.1.2	The importance of consonants . . . . .	112
9.1.3	The importance of word beginnings . . . . .	114
9.1.4	‘Exotic’ phones, suprasegmentals and graphemes . . . . .	115
9.1.5	Lexical stress differences . . . . .	116
9.1.6	Word frequency, neighbourhood density and word length . . . . .	117
9.1.7	Cross-modality influences . . . . .	119
9.1.8	Summary and implications . . . . .	120
9.2	Quantification of predictors . . . . .	122
9.2.1	Formal distance . . . . .	122
9.2.2	Cognate frequencies . . . . .	129
9.2.3	Stimulus length . . . . .	129
9.3	Variable selection . . . . .	129
9.3.1	Bivariate relationships . . . . .	130
9.3.2	Random forest-based conditional permutation importance . . . . .	131

9.3.3	Results and discussion . . . . .	135
9.4	Regression modelling . . . . .	139
9.4.1	Written items . . . . .	139
9.4.2	Spoken items . . . . .	144
9.5	Discussion . . . . .	145
9.5.1	Variables considered . . . . .	147
9.5.2	The use of multiple supplier languages . . . . .	150
9.5.3	Postscript . . . . .	152
<b>10</b>	<b>Participant–item interactions</b>	<b>153</b>
10.1	Possible interactions . . . . .	153
10.1.1	Formal distance and fluid intelligence . . . . .	154
10.1.2	Formal distance and crystallised intelligence . . . . .	154
10.1.3	Cognate frequency and crystallised intelligence . . . . .	154
10.1.4	Interactions with age . . . . .	156
10.2	Method of analysis . . . . .	157
10.3	Interactions between age and item-related predictors . . . . .	158
10.3.1	Written items . . . . .	159
10.3.2	Spoken items . . . . .	163
10.4	Interactions between cognitive and item-related predictors . . . . .	165
10.4.1	Written items . . . . .	165
10.4.2	Spoken items . . . . .	169
10.5	Discussion . . . . .	174
10.5.1	Written items . . . . .	174
10.5.2	Spoken items . . . . .	176
<b>IV</b>	<b>Conclusions</b>	<b>179</b>
<b>11</b>	<b>Synthesis and new directions</b>	<b>181</b>
11.1	Synthesis . . . . .	181
11.2	Avenues for further research . . . . .	183
	<b>Appendices</b>	<b>187</b>
<b>A</b>	<b>Items for the cognate guessing task</b>	<b>189</b>
A.1	Written items . . . . .	189
A.2	Spoken items . . . . .	192

<i>Contents</i>	ix
<b>B Description and results of the Simon task</b>	<b>195</b>
<b>References</b>	<b>199</b>



# Tables

4.1	Main demographic characteristics of the participant sample	50
4.2	Sequence of tasks in the task battery . . . . .	52
5.1	Summary data for the number of correctly translated stimuli per participant in the cognate guessing task . . .	68
5.2	Summary data for the participant-related linguistic and cognitive measures . . . . .	71
6.1	GAMM modelling translation accuracy on written target words in function of age . . . . .	80
6.2	GAMM modelling translation accuracy on spoken target words in function of age . . . . .	81
7.1	GLMM modelling translation accuracy on written target words in function of participant-related predictors . . .	87
7.2	GLMM modelling translation accuracy on spoken target words in function of participant-related predictors . . .	92
7.3	Variable-by-modality interactions . . . . .	94
9.1	Ad-hoc vowel phone inventory . . . . .	124
9.2	Ad-hoc consonant phone inventory . . . . .	125
9.3	Correlations between the item-related predictor variables and the by-item random intercepts . . . . .	131
9.4	Descriptive statistics of the item-related variables affecting written cognate guessing accuracy . . . . .	141
9.5	GLMM modelling translation accuracy on written target words in function of item- and participant-related predictors	142
9.6	Descriptive statistics of the item-related variable affecting spoken cognate guessing accuracy . . . . .	144

9.7	GLMM modelling translation accuracy on spoken target words in function of item- and participant-related predictors	146
10.1	GAMM modelling translation accuracy on written target words in function of interactions between age and item-related predictors . . . . .	160
10.2	GAMM modelling translation accuracy on spoken target words in function of interactions between age and item-related predictors . . . . .	164
10.3	GAMM modelling translation accuracy on written target words in function of interactions between participant- and item-related predictors . . . . .	167
10.4	GAMM modelling translation accuracy on spoken target words in function of interactions between participant- and item-related predictors . . . . .	171
A.1	Written stimuli used in the Swedish cognate guessing task	189
A.2	Spoken stimuli used in the Swedish cognate translation task	192
B.1	Summary data for the Simon task . . . . .	197

# Figures

3.1	Lifespan trajectories of fluid and crystallised intelligence	37
3.2	The multi-component working memory model . . . . .	41
5.1	Target word performance in function of profile word performance . . . . .	69
5.2	Sample age trends in the participant-related predictors .	72
5.3	Intercorrelations between the participant-related predictors	76
6.1	Number of correctly translated target words per participant as a function of age . . . . .	78
6.2	Modelled age trends in translation accuracy . . . . .	82
7.1	Partial fixed effects of the participant-related predictors on translation accuracy (written items) . . . . .	88
7.2	Partial fixed effects of the participant-related predictors on translation accuracy (spoken items) . . . . .	93
8.1	Residual age effects in the random intercepts . . . . .	100
9.1	Example of a Levenshtein distance computation . . . . .	107
9.2	A 7- and an 8-slot Levenshtein alignment for the same cognate pair . . . . .	108
9.3	Ad-hoc vowel phone inventory . . . . .	126
9.4	Illustration of a small random forest . . . . .	133
9.5	Conditional permutation importances (written items) .	136
9.6	Conditional permutation importances (spoken items) . .	137
9.7	Conditional permutation importances with Germanic predictors (written items) . . . . .	140

9.8	Histograms of the item-related variables affecting cognate guessing accuracy for written items . . . . .	141
9.9	Partial fixed effects of the item-related predictors on translation accuracy (written items) . . . . .	143
9.10	Histogram of the item-related variable affecting cognate guessing accuracy for spoken items . . . . .	145
9.11	Partial fixed effects of the item-related predictors on translation accuracy (spoken items) . . . . .	147
10.1	Interaction between age and Germanic Levenshtein distance (written items) . . . . .	161
10.2	Interaction between age and Germanic cognate frequency (written items) . . . . .	161
10.3	Interaction between age and German Levenshtein distance (spoken items) . . . . .	165
10.4	Interaction between crystallised intelligence and Germanic Levenshtein distance (written items) . . . . .	168
10.5	Interaction between fluid intelligence and Germanic Levenshtein distance (written items) . . . . .	169
10.6	Interaction between crystallised intelligence and Germanic cognate frequency (written items) . . . . .	170
10.7	Interaction between crystallised intelligence and German Levenshtein distance (spoken items) . . . . .	172
10.8	Interaction between fluid intelligence and German Levenshtein distance (spoken items) . . . . .	173
B.1	Lifespan development of the Simon effect . . . . .	197

# Preface

In the course of the last three years, I have had the chance to dig into a topic that has fascinated me ever since I started my Bachelor's: how people can understand language varieties that they have never learnt or acquired. This opportunity was made possible by a Sinergia research grant awarded by the Swiss National Science Foundation to Raphael Berthele (project 130457, *Multilingualism through the lifespan*) and by private financial support provided by Dr Ambros Boner, whom I thank sincerely.

I am grateful to my supervisor, Raphael Berthele, for his scientific insights as well as the hospitality and generosity that he and his family continue to extend to me. I also thank my colleagues on the *Multilingualism through the lifespan* project for their work and input, with a special mention of Irmi Kaiser, Lenny Bugayong and Nuria Ristin-Kaufmann, who were instrumental in the data collection stage. Additionally, I would like to thank the other colleagues and staff at the Department of Multilingualism for their congeniality.

Thanks are also due to Charlotte Gooskens and her research team at the University of Groningen as well as to Harald Baayen (University of Tübingen) for their feedback, assistance and hospitality. Thanks also to the developers who make available their state-of-the-art software packages for free and who invest a lot of their time in patiently answering inquiries of end-users—this thesis would have looked wholly different without them. All errors remaining are of course my sole responsibility.

I also want to thank all 167 participants who spent two and half to three hours filling out forms, pressing buttons and providing translations as well as the nineteen students from the Department of Special Educa-

tion (University of Fribourg) who served as the pilot guinea pigs for the cognate guessing task.

Lastly, the most special of thanks go to my parents and to Naomi.

The datasets and the computer code used for the analyses are freely available online at <http://dx.doi.org/10.6084/m9.figshare.795286>, allowing interested readers to fully reproduce all results in this thesis or carry out their own analyses. Please note that the software used for the analyses is in constant development, and later program versions may yield slightly different, more accurate results. Should you spot any errors in my code, please let me know at [janvanhove@gmail.com](mailto:janvanhove@gmail.com).

Note: As of February 2014, a condensed version of Part II of this thesis will appear as Vanhove and Berthele (forthcoming, c) and Chapter 9 in Part III forms the backbone of a forthcoming book chapter by Vanhove and Berthele (forthcoming, b). The results of Part II are also discussed from a more applied perspective in a paper to be published as Berthele and Vanhove (forthcoming). Some of the wording unavoidably overlaps between the two first papers and this thesis.

# Part I

## Introduction



# Chapter 1

## Context and aims

Foreign languages are easier to learn when they show formal similarities to languages that the learner already knows. The effects of formal similarities are manifest in foreign language production and show themselves in the domains of grammar and phonology, but it is particularly in the comprehension of foreign language vocabulary that they are strikingly obvious and overwhelmingly beneficial. For instance, speakers of English without prior competences in Dutch may be able to make some sense of a Dutch text thanks to words that are formally similar to English words, e.g. *chocolade* ‘chocolate’, *wafel* ‘waffle’ or *bier* ‘beer’ (to take a couple of words close to the author’s heart). Such genealogically related words with similar meanings in different languages are known as *cognates*. Due to their genealogical link, they often show a certain degree of formal overlap that learners can capitalise on when listening to or reading a text in the target language. Cognate relationships abound between closely related languages and are sometimes so pervasive and easily identifiable that the languages in question are mutually intelligible.

That said, not all cognates in a given target language are equally easy to identify nor is everyone equally adept at spotting them. In this thesis, I investigate the ability of multilinguals to make educated guesses as to the meaning of words in an unknown language with cognates in one or more languages that they do know. More specifically, I investigate to what extent this ability changes throughout the lifespan and how any age-related changes may be attributed to the participants’ cognitive

and linguistic development. In doing so, I hope to make a contribution to research on the role of cognates in foreign language learning and receptive multilingualism as well as to our understanding of the role of age in learning and using multiple languages. In what follows, I will first flesh out these domains in somewhat greater detail and then discuss the present project.

## 1.1 Cross-linguistic similarities in language learning

When the language learning process is affected by cross-linguistic similarities, this constitutes a case of *cross-linguistic influence*, or *language transfer*, which Odlin (1989) defines as follows:

Transfer is the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired.  
(p. 27)

In the absence of actual cross-linguistic similarities, transfer often leads to errors (negative transfer), but transfer induced by actual cross-linguistic similarities between related languages more often than not has a positive effect, as illustrated by a seminal study by Ringbom (1987). Ringbom (1987) compared the English language skills of Finnish- and Swedish-speaking Finns and found that Swedish speakers outperformed their Finnish-speaking compatriots both in production and in comprehension (see also Ringbom, 2007). Educationally and culturally, both groups are highly similar, leading Ringbom to conclude that the Swedish speakers' advantage must lie in their native language: Swedish is a Germanic language and thus closely related to English. This means that Swedish speakers can make use of a great deal of similarities that exist between Swedish and English, especially in the domains of grammar and lexis, to ease the language learning process. Finnish, by contrast, is a non-Indo-European language and does not offer its native speakers as much in terms of potentially useful transfer bases. As a result, Finnish speakers need to devote considerably more time and effort when learning English than Swedish speakers to reach the same level of proficiency.

Transfer manifests itself in the domains of grammar, phonology, orthography, discourse, pragmatics and sociolinguistic competence (see e.g. Jarvis and Pavlenko, 2008), but its effect is most noticeable and most overwhelmingly positive in the lexical domain, where cognate relationships increase the likelihood that formal similarity goes hand in hand with semantic similarity. For one thing, cognateness makes it easier to memorise foreign language vocabulary that has been imparted explicitly (De Groot and Keijzer, 2000; Lotto and De Groot, 1998). But an arguably far greater benefit of cognate relationships is that readers and listeners who can identify them can tap into a reservoir of potential target language vocabulary without explicit instruction—receptively at first but potentially even productively later on. In general, closely related languages offer the largest potential vocabulary, but even between more distantly related languages such as English and Spanish, cognate relationships may be of use to foreign language learners (Bravo et al., 2007; Dressler et al., 2011; Lubliner and Hiebert, 2011). Capitalising on cognate relationships is consequently often considered an essential part of efficient language learning (e.g. Carton, 1971; Haastrup, 1991; Meißner, 1999; Rubin, 1975).

The presence of cognate relationships does not by itself imply that language learners make use of them, however, i.e. transfer does not automatically take place. The most obvious precondition for cognate relationships to be of use to language learners is that they are aware of them (see Otwinowska-Kasztelanic, 2011, for a discussion). Awareness of cognates is not merely a straightforward function of the target words' formal (objective) similarity to known words: like many other forms of transfer (Kellerman, 1977, 1983), it is primarily governed by how similar the individual language learner *perceives* the cognates to be. This distinction between objective and perceived similarity is likewise relevant in the area of research that I will consider next, namely that of receptive multilingualism and intelligibility patterns between languages.

## 1.2 Receptive multilingualism

Language varieties are sometimes so closely related that they show extensive formal overlap. This can result in speakers of one language (partially) understanding speakers of the other and vice versa without their having learnt or acquired that language and without having to

resort to a common lingua franca: interlingual comprehension takes place as the result of extensive positive transfer thanks to the pervasive cross-linguistic similarities between the languages involved.

The textbook case of such interlingual comprehension is found in Scandinavia between the closely related languages Danish, Norwegian and Swedish. Native speakers, and to a lesser extent non-native speakers, of these languages are able to (partially) understand written and spoken texts in the other languages, even without prior instruction (see e.g. Delsing and Lundin Åkesson, 2005; Haugen, 1966; Maurud, 1976). While the Scandinavian situation stands out in that such interlingual comprehension is often associated with a sense of pan-Scandinavian identity and is actively encouraged by the respective governments, interlingual comprehension also occurs to a greater or lesser extent between other pairs of related languages. Examples include, but are not at all limited to, Afrikaans–Dutch (Gooskens and Van Bezooijen, 2006; Van Bezooijen and Gooskens, 2005a), Czech–Slovak (Berger, 2003), Dutch–German (Gooskens et al., 2011; Ház, 2005) and Portuguese–Spanish (Jensen, 1989). In fact, since the distinction between languages and dialects is arbitrary from a purely linguistic point of view, cases of interdialectal comprehension within a language area can be listed as examples as well (e.g. Chaoju and Van Heuven, 2009; Impe et al., 2008). The phenomenon of interlingual comprehension on the basis of cross-linguistic similarities is variously referred to as *semicommunication*, *intercomprehension*, *mutual intelligibility* or *comprehensibility*, *receptive bilingualism* and *receptive multilingualism*. While each term has a somewhat different denotation and connotation, the differences between them are all in all minimal and I will use the overarching term *receptive multilingualism* (Braunmüller and Zeevaert, 2001) to refer to all of them.

Receptive multilingualism is usually far from perfect in the sense that comprehension is not achieved as easily as it is when listening to or reading in the native language. Often, in fact, mere typological relatedness between two languages is not a sufficient precondition for receptive multilingualism to take place, and a series of learning materials has been developed that aim to foster receptive multilingualism (particularly in the written modality) by imparting cross-lingual reading strategies and increasing awareness of cross-linguistic similarities (e.g. Hufeisen and Marx, 2007; Klein and Stegmann, 2000; Müller et al., 2009; Schmidely et al., 2001). Additionally, even in the absence of such sensitisation,

receptive multilingualism sometimes exhibits a feature that may appear to be rather curious at first sight: it may be skewed, i.e. speakers of one language may understand texts in the other language better than the other way around (e.g. Delsing and Lundin Åkesson, 2005; Gooskens and Van Bezooijen, 2006). While the existence of learning tools for improving receptive skills in a typologically closely related language illustrates that readers and listeners need to be aware of cross-linguistic similarities if they are to make use of them, the asymmetry patterns underscore that awareness is not a straightforward function of the degree of similarity—otherwise comprehension would be as symmetrical as the linguistic distances between the languages. Rather, what is of importance is the perceived degree of similarity, as outlined in the previous section, which can be considered to be a function of objective linguistic factors as well as reader- or listener-related characteristics. Consequently, a considerable body of research has been devoted to identifying the factors affecting success in receptive multilingualism.

Much of this research in the field of receptive multilingualism focusses on the cross-linguistic similarities provided by cognate pairs and therefore on the comprehension of individual words. On the one hand, the rationale is that word comprehension is the single most important factor in receptive multilingualism (Möller and Zeevaert, 2010; Van Heuven, 2008). In the words of Van Heuven (2008),

[t]he underlying assumption here is that word recognition is the key to speech understanding. As long as the listener correctly recognises words, he will be able to piece the speaker's message together. (p. 43)

Morphosyntactic differences between closely related languages are generally small, and their impact on receptive multilingualism is fairly modest in comparison to formal obscurities in cognate relationships, as shown experimentally by Hilton et al. (2013a,b) in the Scandinavian context. On the other hand, investigating readers' or listeners' comprehension of individual words makes it easier to zoom in on the importance of, for instance, specific linguistic features by excluding the possibility that participants make use of co- or contextual cues.

Several studies have consequently made use of what I will call *cognate guessing tasks*. These are tasks in which participants are presented with isolated written or spoken words in an unknown language ( $L_x$ ) that

have cognates in one or more languages that they do know ( $L_1$ ,  $L_2$ ,  $\dots$ ,  $L_n$ ). With no co- or contextual cues to go by, participants need to rely on the cross-linguistic similarities offered by cognate relationships in order to make an educated guess as to the meaning of these words (*interlingual inferencing*; Carton, 1971). Needless to say, performance on cognate guessing tasks varies between participants as well as between stimuli, and a number of participant- and stimulus-related characteristics have been found or hypothesised to affect cognate guessing success. This thesis aims to contribute to the understanding of such factors by considering systematically the role of readers' and listeners' age and age-related factors on cognate guessing success—an aspiration inspired by the importance attached to the age factor in research on language learning and multilingualism. The role of age in multilingualism is discussed briefly in the next section.

### 1.3 Multilingualism and the age factor

The question of how people's age interacts with individual multilingualism is a classic topic in second language acquisition and multilingualism research (see e.g. Singleton and Ryan, 2004, for an overview). On the one hand, researchers have investigated how multilingualism affects the rate of cognitive ageing (e.g. Bialystok et al., 2004, 2005b; Chertkow et al., 2010); on the other hand, a large body of research is concerned with establishing to what extent people's language learning ability and language use is affected by their age. Much of the latter research is specifically devoted to determining if and how the age at which people start to learn a second or foreign language constrains their eventual proficiency in that language. Such studies include those concerned with the *critical period hypothesis* (CPH), which in its most general version holds that learners' susceptibility to naturalistic L2 input is a non-linear function of age (see e.g. Birdsong, 2006; DeKeyser, 2012; Vanhove, 2013, for overviews), as well as those investigating the effects of early foreign language teaching in school contexts (e.g. Muñoz, 2006). The focus in both kinds of studies lies firmly on relatively young learners (but see Hakuta et al., 2003, and DeKeyser et al., 2010, for CPH-inspired studies that include older learners, and Berndt, 2003, and Brändle, 1986, for considerations of older foreign language students) so that, all in all, little is known about how older people cope with new languages.

At the same time, there is hardly any research on how the ability of language users to put cross-linguistic similarities to good use changes as a function of their age. A handful of studies that will be discussed in more detail in Section 2.4 (viz. Berthele, 2011; Cenoz, 2001; Delsing and Lundin Åkesson, 2005; Schüppert et al., forthcoming) suggest (directly or indirectly) that this ability does not remain constant throughout the lifespan, but its precise age-patterning throughout the lifespan remains unknown as do the cognitive and linguistic driving forces behind such age-related development. With this thesis, I aim to shed light on precisely these issues by investigating how the cognate guessing task performance of multilingual participants aged 10 to 86 years changes as a function of their age and to what extent this development can be ascribed to age-related cognitive and linguistic changes. The research presented in this thesis was conducted as part of a larger collaboration that I will briefly present next.

## 1.4 The present project

### 1.4.1 The overarching project ‘Multilingualism through the lifespan’

This thesis was written under the auspices of the project *Multilingualism through the lifespan* funded through the Sinergia programme of the Swiss National Science Foundation (SNF-130457, PI Raphael Berthele). This project’s goal is to examine the impact of ageing and concomitant cognitive and social changes on various aspects of multilingualism. It comprises six subprojects, three of which are sociolinguistically and ethnographically oriented and three of which have a more psycholinguistic slant to them. The three latter subprojects, among which my project, rely on the same pool of multilingual Swiss participants aged 10 to 86 years with Swiss German and standard German as their L1s,<sup>1</sup> allowing for potentially meaningful comparisons across subprojects.

---

<sup>1</sup>How best to describe the linguistic situation in the German-speaking part of Switzerland is somewhat of a vexed question. Some authors consider it a diglossic situation in which speakers have both their dialects and the standard language as first languages, whereas others argue that the standard language is a second language to most speakers (see Berthele, 2004, for more discussion). The issue is noted here, but it is of no further consequence to this thesis.

I will not describe the other five subprojects in any detail as their eventual overall synthesis is beyond the scope of this thesis; for a detailed overview, I refer to the upcoming Vol. 99, Issue 1 of the *Bulletin suisse de linguistique appliquée* (projected for June 2014), which is devoted to the Sinergia project. Briefly, however, what binds all subprojects together apart from a shared interest in the age factor in multilingualism is that they define multilingualism very broadly: it is not the ‘perfect’, balanced multilinguals that take centre stage, but rather the language users who, confronted with multilingual variation, have to ‘make do’ and marshal their cognitive and linguistic resources to cope with the challenge at hand. Nor are these projects concerned with one of the mainstays of linguistic research into the age factor, the question of the ‘ultimate attainment’ states, i.e. what second and foreign language learners are maximally capable of after years of L2/ $L_n$  exposure. In fact, the focus in two subprojects (including mine) is on the very first steps in a new language.

## 1.4.2 Aim, scope and terminology

### Aim

I pursue three aims in this thesis. First, given the dearth of research on age-related developments in language transfer in general and receptive multilingualism in specific, I investigate how the performance on  $L_x$  cognate guessing tasks develops throughout the lifespan in multilingual participants. In order to address this question, I administered a cognate guessing task featuring a total of 100 Swedish words (50 written, 50 spoken) to a sample of 163 multilingual participants aged 10 to 86 years with Swiss German and standard German as their L1s and French and English as common second or foreign languages. Ninety of these Swedish words had German, English or French translation-equivalent cognates and their meaning could therefore theoretically be inferred by the participants.

Second, as people grow older, they improve in certain linguistic and cognitive domains and become worse in others, as I will discuss in Chapter 3. The next question, therefore, is whether and how any age trends found with respect to cognate guessing task performance can be attributed to age-related improvements and declines in such cognitive and linguistic domains. To answer this question, language proficiency

data as well as a selection of cognitive variables were collected for all participants.

Lastly, much research on receptive multilingualism has shown how accurate cognate guessing is dependent on certain properties of the cognates that are presented. With increasing formal overlap between the stimuli and their known cognates, for example, accurate cognate guessing becomes more likely. It is conceivable, however, that the precise effects of such item-related characteristics vary somewhat from participant to participant. This in turn gives rise to the possibility that the participants' susceptibility to these item-related characteristics is systematically related to their linguistic and cognitive resources and hence ultimately to their age, too. The third research question of this thesis is whether this is indeed the case.

### **Delineating the scope**

In addition to specifying what this thesis is about, I think it is useful to make explicit what it is *not* about. First, while I am highly indebted to research on mutual intelligibility patterns between language varieties, what interests me here is not the mutual intelligibility between (Swiss) German and Swedish. The choice of Swedish as the target language is, in fact, rather incidental, and stimuli from another Germanic language would likely have yielded results similar to the ones presented in this thesis (see Vanhove and Berthele, forthcoming, a, for how results gained from cognate guessing tasks generalise to other related languages).

Second, I am concerned strictly with the ability to understand isolated words, and the results discussed here cannot be directly extrapolated to the comprehension of full sentences and full texts in a related foreign language: while I assume that the ability to understand single cognate words is fundamental to full sentence and full text comprehension (Möller and Zeevaert, 2010; Van Heuven, 2008), this ability interacts with the ability to make inferences based on co- and contextual cues. From an applied perspective particularly, my focus may therefore strike the reader as narrow. In my opinion, however, this narrow scope is justified by the desire to get to the bottom of one type of inferencing (interlingual) by reducing the role of other types of inferencing (intra-lingual, extra-lingual) to a minimum.

This brings me to a third point. The Sinergia project as a whole is intended as a contribution to basic research on multilingualism. I will

therefore not discuss the applied perspectives that the results presented in this thesis may offer. However, some applied implications are discussed by Berthele and Vanhove (forthcoming).

Lastly, mentioning ‘cognates’ and ‘multilinguals’ in the same sentence tends to evoke associations with research on lexical access in bi- and multilinguals that exploit the properties of cognates (e.g. Costa et al., 2000; Dijkstra and Van Heuven, 2002; Lemhöfer and Dijkstra, 2004). Interesting though such studies are, they are concerned with the automatic activation of lexical items in the bi- or multilingual brain and make use of bi- or multilingual participants that have extensive prior knowledge of the target language. In cognate guessing tasks, by contrast, participants have no prior knowledge of the language that they are confronted with, and whatever automatic lexical activation effects the stimuli may bring about are buried under a thick layer of (often effortful) decision processes.<sup>2</sup>

### Defining cognates

The central term in this thesis, i.e. *cognate*, does not have a universally agreed upon definition (see also Berthele, 2011). Whereas Dijkstra et al. (1999), for instance, define cognates very strictly in terms of formal and semantic overlap (“[t]hose interlingual homographs that not only share their orthographic form but their semantics as well”, p. 497) and Lotto and De Groot (1998) use a slightly more inclusive but vaguer definition that still appeals to the word forms (“[w]ords with similarly formed translation equivalents in the L2”, p. 38), Kürschner et al. (2008) define them in terms of their historical link and semantic overlap (“historically related word pairs that still bear the same meaning in both languages”, p. 86). It is necessary, therefore, to specify how I intend the word: I define cognates as historically related words in different language varieties that are translation equivalents in at least one sense. Thanks to their historical link, cognate pairs often show some formal resemblance, but the degree of formal overlap varies from cognate pair to cognate pair and formal resemblance is neither a necessary nor a sufficient condition to establish cognateness. My definition is intended to be maximally inclusive and to capture word pairs tracing back to the same root in an

---

<sup>2</sup>That said, studies such as the ones cited are still relevant for my purposes as they underscore that the languages in a multilingual repertoire are intimately linked and interact with each other, even unbeknownst to the participant.

ancestor language (*Erbwörter*, e.g. Dutch *huis* and English *house*) and loan words (e.g. English *keelhaul* from Dutch *kielhalen*, but also loan words borrowed from a third language, e.g. Dutch and English *sauna* from Finnish), as well as calques formed according to the same pattern (e.g. Dutch *invloed*, from Latin *influxus*, and English *influence*, from Old French).

## 1.5 Overview

This thesis is broadly organised along the research questions outlined above. First, Part II focusses on participant-related effects in cognate guessing, especially as they relate to age-related developments. It starts with an overview of participant-related factors that have already been investigated in studies on receptive multilingualism and cognate guessing in specific (Chapter 2). Then, Chapter 3 turns to research on cognitive age-related developments and discusses its potential relevance for cognate guessing. Chapters 4 to 8 discuss the design and results of an empirical investigation into the participant-related factors discussed before.

Part III aims to complement the findings of Part II by considering item-related effects on cognate guessing in addition to participant-related effects. In Chapter 9, I discuss various item-related factors that may be at play in cognate guessing and verify whether they do indeed affect the participants' performance in the present study. In Chapter 10, I investigate whether these item-related factors have varying effects depending on the participants' age and cognitive characteristics.

Part IV synthesises the findings of this thesis and offers some perspectives for further investigation.



## Part II

# The lifespan development of cognate guessing skills



## Chapter 2

# Inter-individual differences in cognate guessing skills

The present part of this thesis is concerned with inter-individual differences in cognate guessing skills. In one form or another, such inter-individual differences have been the subject of a host of studies on receptive multilingualism. The thematic novelty of my contribution to this topic lies firstly in the fact that I consider the effect of the participants' age much more systematically than has thus far been the case. Second, this is—to my knowledge—the first study that investigates the role of the cognitive factors intelligence and working memory on cognate guessing. These cognitive factors are discussed in Chapter 3. First, however, I review factors which in previous studies have been found to give rise to inter-individual differences in the aptitude for receptive multilingualism and cognate guessing skills more specifically and discuss the implications for the present study.

## 2.1 Linguistic repertoire

### 2.1.1 Typological relation between the $Lx$ and the L1

In order to be able to partially understand an unknown language  $Lx$ , it is evidently necessary to have linguistic resources in one's repertoire that provide a 'bridge' towards the  $Lx$ . For most purposes, such potentially useful linguistic resources are languages that are typologically closely related to the  $Lx$  as these provide the most cross-linguistic similarities in terms of morphology, syntax and, of course, lexis. That said, two caveats are in order. First, languages more distantly related or indeed unrelated to the  $Lx$  may, in principle, provide useful comprehension cues in relatively isolated instances. Knowledge of French, for instance, may help an L1 speaker of German to understand certain words in the non-Romance language Dutch that do not have a cognate in German, e.g. *punaise* 'drawing pin'. Second, the literature on transfer in foreign language learning stresses that what governs the likelihood of transfer is not so much the actual (objective) cross-linguistic similarities and typological relationship that exist between two languages but rather whether and how these similarities and this relationship are perceived by the participants (*perceived similarities* and *psychotypology*, respectively; Kellerman, 1977, 1983; see also De Angelis, 2007, pp. 22–33; Jarvis and Pavlenko, 2008, pp. 176–182; Ringbom and Jarvis, 2009).

All in all, however, closely related languages offer the greatest potential for perceiving similarities, and a high degree of proficiency in a closely related language has been found to be beneficial to  $Lx$  comprehension (see also Meißner and Burk, 2001). Swedish-speaking Finns, for instance, unsurprisingly outperform their Finnish-speaking compatriots, for whom Swedish is a school subject, in Danish and Norwegian reading and listening comprehension (Delsing and Lundin Åkesson, 2005). A related finding is that Swedish speakers in Finland generally score better than Finnish speakers on nationwide English exams (Ringbom, 1987), which has been attributed to, among other things, their head start in vocabulary comprehension thanks to the widespread cognate relationships between English and Swedish (see Ringbom, 2007, p. 11). Other things being equal, then, having a 'bridging language' as one's L1 is a particularly advantageous precondition for  $Lx$  comprehension.

Furthermore, the usefulness of such a bridging language is a function of its linguistic distance to the *Lx*: generally speaking, the closer the two languages are linguistically speaking, the better comprehension between them will be. This is illustrated by several studies by Gooskens and colleagues, who have shown phonetic distance to be a reliable predictor of spoken *Lx* text comprehension (Beijering et al., 2008; Gooskens, 2007a,b; Gooskens et al., 2008), and non-correlational findings point to a similar role of orthographic distance in written *Lx* text comprehension (Van Bezooijen and Gooskens, 2005a,b). For the comprehension of individual written and spoken words, similar findings have been reported, which will be discussed in Chapter 9 in Part III.

In the present study, all participants share the same bridging languages as L1s, viz. standard German and a Swiss German dialect (see Note 1 on page 9). While Swiss German is an umbrella term covering the different Alemannic dialects spoken in Switzerland, my working assumption is that the contribution of any differences in the linguistic and average psychotypological distances between different Swiss Germanic dialects and the *Lx* in question (Swedish) to inter-individual differences in *Lx* comprehension is negligible. I therefore assume my participant sample to be effectively homogeneous with respect to the L1 bridging language. But even within a population with the same bridging languages as L1s, inter-individual differences in L1 experience and breadth of vocabulary may give rise to inter-individual variance in *Lx* comprehension. I postpone the discussion of this potential source of *Lx* comprehension variance to Section 3.1 on page 34, where it will be subsumed under the label of ‘crystallised intelligence’. Moreover, L1 knowledge of one or more bridging languages does obviously not preclude readers and listeners from also drawing on other language varieties in their linguistic repertoires as potential bridges towards understanding the *Lx*, an issue which I will turn to next.

## 2.1.2 The impact of multilingualism

### Empirical findings

Inspired by research showing that knowledge of a foreign language facilitates the learning of a new one (e.g. Cenoz et al., 2001), Gibson and Hufeisen (2003) set out to investigate how 36 multilingual language students at a German university went about making sense of a simple

written text in an unknown language, Swedish. In a subsample of 26 learners of German as a foreign language, those who had learnt German as their L4 (i.e. chronologically the third foreign language;  $n = 9$ ) outperformed those who had learnt German as their L2 ( $n = 7$ ), with those having learnt German as an L3 ( $n = 10$ ) falling in-between. Despite the authors' claim to the contrary (p. 102), however, these results do not by themselves indicate that  $Lx$  comprehension is facilitated by the sheer number of languages in one's repertoire, as the L2/L3/L4 status of German was not strongly correlated with the number of foreign languages in the participants' repertoires (cf. their description on p. 99).

A more direct assessment of the effect of previous foreign language skills on receptive multilingualism generally and cognate guessing specifically was first carried out by Berthele and Lambelet (2009), who investigated the performance of 140 Italian- or French-speaking participants on a cognate guessing task featuring a total of 29 isolated written words in the Romance languages Romansh and Romanian. Berthele and Lambelet (2009) found a medium-sized positive correlation between the number of languages in the participants' repertoires and their success on the cognate guessing task ( $r_s = 0.35$ ). Revisiting data from an earlier study involving 179 German-speaking participants decoding 29 isolated written Scandinavian words (Berthele, 2008), Berthele (2011) similarly reports a medium positive effect of the size of the linguistic repertoire ( $r = 0.25$ ).

By themselves, these results may appear to suggest that knowledge of several languages is somewhat conducive to the comprehension of written isolated  $Lx$  stimuli. However, Berthele's (2011, Table 2) stepwise multiple regression model suggested that the number of languages in the repertoire was actually a *negative* predictor of how well 163 Swiss German participants could understand 28 written and spoken isolated Scandinavian words ( $B = -0.24 \pm 0.11$ ). Furthermore, two studies by Marx on written  $Lx$  text comprehension in German students (Marx, 2007, 2011) did not reveal any association between the participants' performance on the one hand and the number of foreign languages they knew or their self-assessed proficiency in these languages on the other hand. All in all, then, the sheer number of languages in the participants' repertoires does not appear to be too useful a predictor when modelling inter-individual differences in cognate guessing skills.

What appears to be rather more crucial is whether participants have good to excellent competences in two or more languages relatively closely

related to one another as well as to the  $Lx$ . On a macro-level (i.e. on the level of language communities), Gooskens (2007b) found that Frisians, who are bilingual in the closely related languages Dutch and Frisian, performed slightly better on a text comprehension task with spoken Afrikaans, another West-Germanic language, compared to non-Frisian Dutch participants (a 4.2 percentage point (pp) difference), whereas Swedish-speaking Finns with excellent competences in the non-Indo-European language Finnish actually understood Danish and Norwegian spoken texts worse than did Swedes (see also Delsing and Lundin Åkesson, 2005). On the level of the individual participants, Singleton and Little (1984) had already noted that English-speaking students with knowledge of German were better able to make sense of a text in another Germanic language, Dutch, than English-speaking students without knowledge of German. With regard to the comprehension of isolated words, Berthele and Lambelet (2009) found that natively Romance-speaking participants with self-reported above-average competences in one additional Romance language understood about 10 pp more written Romansh and Romanian words than Romance-speaking participants with above-average competences in another, non-Romance language. Additionally, Berthele's (2011) stepwise regression model revealed that self-assessed skills in an additional Germanic language, viz. English, are positively associated with Swiss Germans' comprehension of isolated spoken Danish and Swedish words ( $B = 0.31 \pm 0.11$ ). (Note that the positive correlations between the total number of languages, be they related to the  $Lx$  or not, and written  $Lx$  stimulus comprehension found by Berthele and Lambelet (2009) and Berthele (2011) may well be largely by-products of a more important correlation involving the number of languages related to the  $Lx$  in their participants' repertoires.)

That said, not all studies found such a reliable advantage in receptive multilingualism and cognate guessing for participants that are bi- or multilingual in language varieties that are closely related to one another and to the  $Lx$ . Specifically, Van Bezooijen et al.'s (2012) Frisian–Dutch bilinguals ( $n = 39$ ) did not convincingly outperform Dutch speakers without competences in Frisian ( $n = 33$ ) on a listening comprehension task with 384 isolated Danish words, the effect size being merely 0.3 pp. Marx (2007, 2011) similarly found no effect of knowledge of an additional Germanic language on written  $Lx$  text comprehension in German

students, but with cell sizes of six and eight participants, the statistical power to detect even large effects was small.

The studies cited above all investigated the effect of knowledge of more than one standard language on receptive multilingualism, but the effect of dialect knowledge has also been investigated. Berthele (2008) reports positive effects of self-reported receptive and productive competences in at least one German dialect in addition to the standard language on understanding 29 written Scandinavian words presented out of context (a 4 to 7 pp difference). Further evidence that knowledge of an additional dialect closely related to the *Lx* may enhance spoken *Lx* comprehension is reported by Gooskens et al. (2011). These authors found that Dutch participants hailing from the Dutch–German border region outperformed participants from other parts of the Netherlands in translating spoken isolated Low German words by some 6 pp. The dialects spoken in this border region, like Low German but unlike standard Dutch, belong to the Low Saxon dialect group. This suggests that participants from the border regions make use of both standard language and dialectal representations when trying to make sense of *Lx* stimuli.<sup>3</sup> Lastly, in the Scandinavian context, Norwegians’ knack for understanding Danish and Swedish (see Delsing and Lundin Åkesson, 2005) may be partly attributable to the strong role of dialects in Norwegian society (see e.g. Gooskens, 2007b, but see also Gooskens and Heeringa, forthcoming, for counter-evidence).

Summarising, the general picture emerging from research on the effect of the participants’ multilingual repertoire on their aptitude for understanding *Lx* stimuli is that multilingualism plain and simple does not necessarily confer an advantage in this respect. Rather, it is a highly specific form of multilingualism that appears to be more reliably beneficial, viz. being proficiently multilingual in varieties that are closely related to one another and to the *Lx*.

---

<sup>3</sup>Interestingly, these authors also found indications that even the participants from the border region gave precedence to standard Dutch as a bridging language and tended to rely on their dialect knowledge mainly when standard Dutch was of no help. Speculatively, these dialects are sociolinguistically marked and participants may not consider them as equally suitable transfer sources as the standard as a result (see James, 1983).

### Possible explanations of a multilingualism effect

The advantage that participants with high competences in two or more language varieties closely related to an  $Lx$  may enjoy in decoding isolated  $Lx$  stimuli may stem from their simply having more potentially helpful transfer bases at their disposal. Indeed, Swarte et al. (2013) showed that the advantage that Dutch participants with high proficiency in German have in understanding isolated written Danish words relative to Dutch participants with lower German proficiency levels is due mainly to their better comprehension of Danish–German cognates. It is for this reason that Van Bezooijen et al. (2012) expressed some surprise at their null result, as their Frisian–Dutch bilinguals might have been expected to approach the  $Lx$ , Danish, via one additional Germanic language compared to non-Frisian Dutch subjects and therefore have been at an advantage.

An additional possible explanation lies in the heightened metalinguistic skills that emerge from the interplay between language varieties in the multilingual mind, the so-called ‘M-factor’ (Herdina and Jessner, 2002). Consider what is required of participants in cognate guessing tasks. They are confronted with  $Lx$  stimuli that in most cases do not completely match the corresponding entries in their  $L1$ ,  $L2$ , ...,  $L_n$  vocabularies. Consequently, they need to be somewhat flexible vis-à-vis formal discrepancies between the  $Lx$  stimuli and their potential  $L1$ ,  $L2$ , ...,  $L_n$  translation equivalents (*Wahrnehmungstoleranz*; Berthele, 2008). If participants are only slightly tolerable of formal discrepancies, they will only provide answers to the fairly obvious cases (e.g. internationalisms or other form-identical cognates) and decline to respond in cases where no obvious translation equivalent presents itself. But in order to maximise their performance on the task (and assuming they would want to), participants need to take risks and make guesses in cases that are not so obvious. With increased tolerance levels come larger search spaces, but participants cannot rely on co- and contextual cues to efficiently limit these search spaces.<sup>4</sup> The only top-down processes which they can apply are based on their subjective estimates of the probabilities with which particular cross-linguistic correspondences occur (“linguistisches Probabilitätskalkül”; Berthele, 2008, p. 92).

---

<sup>4</sup>In receptive multilingualism tasks using full sentences or texts, participants obviously can use such cues to limit their search spaces.

How might participants know which potential cross-linguistic correspondences are likely and which are unlikely? A possible mechanism is discussed by Berthele (2011), who proposes that multilinguals in related languages may develop a meta-system by way of abstracting away from differences between these languages and instead focussing on cross-linguistic similarities.  $Lx$  stimuli may then not only be linked to their possible L1, L2, . . . ,  $L_n$  cognates but also to their abstract meta-forms. These meta-forms, by their very nature, have been proven to be more robust cross-linguistically and may therefore give rise to more fruitful decoding attempts. Thus, Berthele (2011) argues, French-speaking participants with knowledge of Italian may be in a superior position compared to monolinguals when decoding the Romansh word *cuorer* ‘to run’. The corresponding L1 and L2 cognates are *courir* and *correre*, respectively, but French–Italian bilinguals may be able to abstract away from these language-specific forms by focussing on L1–L2 similarities. In this case, such an abstraction may consist of a fully retained consonantal skeleton that maps perfectly onto the  $Lx$  consonantal skeleton.

Importantly, the effect need not be restricted to cases in which the  $Lx$  stimulus has two or more cognates in the multilingual’s lexicon. Multilinguals, Berthele (2011) proposes, may also distil from their multilingual repertoire certain patterns for which they can posit tentative rules. An English–German bilingual, for instance, may observe that German post-vocalic [f] often corresponds to English [p] in cognate pairs such as *tief*–*deep* or *Schiff*–*ship*. A tentative rule for this pattern could be: “Post-vocalic English [p] tends to map onto post-vocalic German [f].” The domain of such a tentative rule may subsequently be enlarged so as to include new  $Lx$  cases in a process named *abduction* (after Peirce, 1934). Thus, the English–German bilingual may speculate about German–Dutch correspondences as well, for instance: “Post-vocalic Dutch [p] might similarly be the counterpart of post-vocalic German [f].” When confronted with the Dutch word *dorp* ‘village’, she may consequently (in this case correctly) infer that it is a cognate of German *Dorf*. In similar vein, the rule itself may be generalised. In the present example, the bilingual could hypothesise that Dutch [p] might map onto German [f] regardless of its position (which is not generally true), or that post-vocalic plosives in Dutch likely correspond to fricatives with the same place of articulation in German (which is often the case). Pajak (2010) presents evidence

that bilinguals are indeed capable of generalising known phonological rules to different segments, contexts and categories in a novel language.

### Factors attenuating a multilingual advantage

While multilingualism, particularly in closely related languages that are themselves closely related to the  $L_x$ , may affect cognate guessing skills, such an effect is almost necessarily not purely positive. Specifically, knowledge of additional languages increases the likelihood that so-called ‘false friends’ come into the picture, leading to negative transfer. Danes, for instance, often translate the Swedish word *art* ‘sort, kind’ as *kunst* ‘art’, even though the Danish cognate *art* is in fact the correct translation (Kürschner et al., 2008), undoubtedly due to their knowledge of English. More generally, if an  $L_x$  word with many L1 neighbours (i.e. words similar in form) is more difficult to infer than one with few neighbours (see Section 9.1.6), it is plausible that  $L_x$  stimuli with many neighbours in the whole multilingual lexicon are also more difficult to infer than words with few multilingual neighbours. Participants with large multilingual vocabularies may therefore be led astray if they cannot distinguish plausible translation equivalents from unlikely ones.

A second factor is that multilingual participants do not necessarily make use of their multilingual repertoires, even if doing so would be to their advantage. Müller-Lancé (2003) and Ender (2007, pp. 199–200), for instance, found that many participants refrained from drawing on multilingual resources but instead applied knowledge of one language only when reading in a foreign language. Müller-Lancé (2003) refers to such participants as ‘monolinguals’. In this context, it bears repeating that what matters in language transfer is not so much the objective similarity between two languages but rather their perceived similarity and the psychotypical relationship between them (see Section 2.1.1).

### Implications

While this study focusses on participants that share the same L1 (German and Swiss German dialects), differences in their linguistic repertoires may nevertheless give rise to inter-individual differences in cognate guessing skills and need to be taken into account. First, the cognate guessing task at hand featured the North-Germanic language Swedish, and participants with knowledge of other North-Germanic languages

(especially Norwegian or Danish) would be unduly favoured in such a task; all participants with knowledge of any North-Germanic language were therefore excluded from the scope of my investigation. Second, foreign language skills are likely to show an age-related development, mainly due to the effect of schooling. To the extent that knowledge of foreign languages contributes to cognate guessing skills, this may therefore give rise to age trends in cognate guessing skills, too. Foreign language knowledge thus needed to be taken into account as a predictor variable. Third, French and English are compulsory school subjects in Switzerland. While knowledge of both languages can be useful when guessing the meaning of isolated Swedish words, English, as a Germanic language, was especially likely to be considered a more appropriate transfer source by the participants. Knowledge of English was therefore measured using both a language test and a self-assessment form. For lack of time, knowledge of French was only measured using a self-assessment form.

## 2.2 Previous exposure

Intuitively, it is readily recognised that prior experience with a given  $Lx$  is conducive to  $Lx$  comprehension. This commonsensical insight has been borne out by empirical investigations. Jensen (1989), for instance, compared how well Spanish-speaking and Brazilian participants understood audio recordings in Portuguese and Spanish, respectively. The amount of previous contact with the target language turned out to correlate positively with oral text comprehension across both language groups, even after excluding the results of the informants with substantial prior experience with the other language ( $n = 71$ ,  $r = 0.30$ ). In the Scandinavian context, Delsing and Lundin Åkesson (2005, Table 5.6, p. 104) similarly report mostly significant rank correlations between an index of contact patterns with the  $Lxs$  and spoken and written text comprehension in the  $Lxs$ , i.e. the national languages of the participants' neighbouring countries (Denmark, Norway and Sweden).<sup>5</sup>

The causal relationship between such contact patterns and  $Lx$  comprehension is not necessarily as straightforward as it appears to be,

---

<sup>5</sup>Unfortunately, Delsing and Lundin Åkesson (2005) do not report the correlation coefficients themselves but rather their associated  $p$ -values, making it difficult to assess the importance of these contact patterns in  $Lx$  comprehension.

however. Delsing and Lundin Åkesson (2005, p. 145) make the profound observation that the relationship between contact and comprehension may be a feedback loop in which participants who can understand an *Lx* rather well may be more likely to avail themselves of opportunities to come into contact with this *Lx*, resulting in even better *Lx* comprehension. Exposure to an *Lx* may thus simultaneously be both the cause and the effect of good *Lx* comprehension.

From a psycholinguistic point of view, this problem is compounded by the learning process that *Lx* exposure entails. This learning process, which—needless to say—is what makes previous *Lx* exposure beneficent to *Lx* comprehension, is according to Warter (1995, cited in Warter, 2001) characterised by three phases with fuzzy boundaries. First, the correspondence rules by which the *Lx* sound system can be mapped onto a known sound system are acquired (*habituation*). Once these correspondence rules are internalised, they are used to ‘correct’ the acoustic form of *Lx* items to fit the form of known items prior to lexical access (*reconstruction*, see also Bannert, 1981). Finally and after a great deal of *Lx* exposure, the *Lx* items themselves have become internalised and may be accessed directly without the need for acoustic correction (*lexical identification*, see also Bannert, 1981). Thus, when trying to make sense of *Lx* items, inexperienced *Lx* listeners are likely to use a strategy that is qualitatively different from that of experienced *Lx* listeners, for whom the *Lx* is effectively an additional known language as far as receptive competences are concerned. While Bannert (1981) and Warter (1995, 2001) made these distinctions with oral interactions in a Scandinavian setting in mind, I see no reason why they cannot readily be transplanted to non-Scandinavian settings or, *mutatis mutandis*, to the comprehension of written *Lx* items.

Given the difficulties in interpreting contact–comprehension relationships and the possibility of participants using different experience-induced strategies, the present study controls for the participants’ previous *Lx* exposure by only including participants without such previous exposure. In doing so, I follow Van Heuven (2008) in limiting the scope of my research to first-exposure situations.

## 2.3 Attitudes

Since Wolff (1959) noted that intelligibility patterns between Nigerian tribes with closely related languages were a function not only of linguistic proximity but of prestige and attitudes as well, skewed intelligibility patterns elsewhere have often been ascribed to attitudinal factors, too (e.g. Bø, 1978; Maurud, 1976). Concretely, speakers belonging to a less prestigious linguistic group may be more motivated to try to understand the language of the more prestigious group than vice versa, resulting in asymmetric comprehension.

However, while attitudes towards the *Lx*, its speakers or their country may be involved in such naturalistic settings, evidence from more controlled testing situations does not lend unequivocal support to the hypothesis that attitudes and *Lx* comprehension are necessarily intimately tied. Neither Jensen (1989) nor Van Bezooijen and Gooskens (2005b) found any significant correlations between their participants' *Lx* text comprehension and their attitudes towards the languages in question, whereas Delsing and Lundin Åkesson (2005, pp. 108 and 110, Tables 5.10 and 5.12; but see Note 5) and Gooskens and Van Bezooijen (2006) found that only a few of the attitude indices that they extracted correlated significantly with *Lx* text comprehension. This led Gooskens and Van Bezooijen (2006) to conclude that the attitudinal factor plays "at the most a weak role" in *Lx* comprehension (pp. 548–549). In addition, Schüppert and Gooskens (2011) found that the reaction times of 86 Scandinavian preschoolers and young adults to isolated *Lx* stimuli could not be predicted from their attitudes towards the *Lx*.

All in all, it seems unlikely that attitudinal factors with respect to the *Lx* play a large role in *Lx* comprehension in testing situations. In such relatively unnaturalistic settings, these attitudes probably take a backseat to other factors affecting the participants' motivation, such as their willingness to take tests and provide translations (see also Wolff, 1959). A further issue to consider is that, even in cases where attitudes could be shown to be reliably related to *Lx* comprehension, it is not clear how such correlations should be interpreted in any causal sense. As Schüppert and Gooskens (2011) point out, participants with positive attitudes may encourage themselves to make a greater effort in *Lx* comprehension, but good *Lx* comprehension may in itself also result in a more positive attitude. Thus, as with the contact factor, the

link between attitudes and comprehension may also be a feedback loop. Schüppert and Gooskens even recognise the possibility that attitudes may impact comprehension indirectly by co-determining the amount of contact with the *Lx*. I add that, by the same token, the amount of contact may also indirectly influence attitude by first leading to better *Lx* comprehension.

Importantly, all studies cited in this section investigated the comprehension between geographically close and contact-intensive language communities (with the exception of Afrikaans–Dutch) yet could at best establish a weak impact of attitudes on *Lx* comprehension. Since the amount of previous *Lx* exposure at the individual level was eliminated as a factor in the present study (see Section 2.2), I consider it even less likely that attitudes come into play in the cognate guessing task used in this study. Consequently, I did not consider *Lx*-related attitudes as predictor variables in this study.

## 2.4 Age

Despite large bodies of research on language transfer and on the age factor in language learning and multilingualism, the relationship between age and language transfer is largely virgin territory. To my knowledge, the first study to explicitly target this relationship was conducted by Cenoz (2001). Cenoz (2001) investigated how the transfer tendencies of 90 Basque–Spanish bilingual learners of English aged 7 to 14 changed as a function of their age. She found, among other things, that the older learners were more likely to transfer Spanish (as opposed to Basque) elements when speaking English than younger learners. Since Spanish, like English, is an Indo-European language, whereas Basque is completely unrelated to English, transferring Spanish elements can be regarded as typologically more sensible than transferring Basque elements, and Cenoz (2001) suspects that this age trend is due to older learners being more cognitively mature and having higher levels of meta-linguistic awareness.

A finding more directly pertaining to the age factor in the comprehension of closely related languages stems from Delsing and Lundin Åkesson's (2005) large-scale study in Scandinavia. In order to determine whether Scandinavians had become better or worse at understanding their respective so-called 'neighbouring languages' in the course of the last few decades, Delsing and Lundin Åkesson (2005) compared adoles-

cents' results on *Lx* comprehension tasks with those of their parents. They found that the parents understood written and spoken texts in the neighbouring languages better than did their teenage children. The difference was particularly pronounced for spoken text comprehension. The authors advanced personal development (more specifically a larger L1 vocabulary and more experience in decoding the specific *Lx*s and in coping with linguistic variation), societal changes (which possibly had caused Scandinavians to be less oriented towards Scandinavia and more towards the European continent and the world at large) and recent linguistic changes (which had caused the Scandinavian languages to drift somewhat further apart) as possible causes for this finding (Delsing and Lundin Åkesson, 2005, pp. 142–144). Unfortunately, their test format did not allow them to tease these factors apart.

Further indications that age is linked to *Lx* comprehension, and to the comprehension of single *Lx* words in particular, are presented by Schüppert et al. (forthcoming). 116 Danish and Swedish 7- to 16-year-olds participated in an auditory word recognition task in the respective other language. In this task, an *Lx* word was played to the participants who then had to point to the corresponding picture out of a possible four presented on the computer screen. Task performance was strongly correlated with the participants' age ( $r = 0.61$ ), especially when taking into account that the participants aged 12 years and older performed roughly at ceiling.

Lastly, in a study resembling mine more closely, Berthele (2011) asked Swiss German participants aged 13 to 35 to translate written and spoken Swedish and Danish words presented in isolation into German. After carrying out a stepwise regression analysis, he found that the participants' age was the most important predictor of *Lx* comprehension, explaining the bulk of the inter-individual variance in comprehension scores eventually accounted for (34% out of 62% in total). In contrast to Delsing and Lundin Åkesson's (2005) study, the effect of societal and linguistic changes affecting the comprehension of Scandinavian words by the participants can reasonably be assumed to be non-existent in Berthele's (2011) study, bringing into focus patterns of personal development.

It is precisely these age-related patterns of personal development that lie at the heart of this study. It is reasonable to hypothesise that growth of L1 as well as foreign language vocabulary, increased experience in

dealing with linguistically deviating varieties even in the L1 and increased meta-linguistic awareness resulting from the interplay between these two factors reflect positively on the ability to decode stimuli in an unknown but related language variety. However, it is important to recognise that ageing is a process that affects multiple facets of cognitive functioning, which in turn may give rise to age differences in *Lx* decoding abilities. Such age-related changes in cognition and their possible repercussions on *Lx* cognate guessing form the subject of the next chapter.



## Chapter 3

# The lifespan development of cognition

I concluded the last chapter by discussing how cognate guessing skills have been found to change as a function of age. Some of the authors cited already advanced the participants' cognitive as well as linguistic development as a factor leading to better cognate guessing skills, but a direct investigation of the link between cognitive development and cognate guessing skills has not yet been undertaken. Furthermore, studies on cognate guessing have hitherto made use of relatively young participants. Cognitive abilities, however, change throughout the entire lifespan, raising the question of how such cognitive changes affect cognate guessing across a broader age range.

This chapter introduces three cognitive constructs that show well-known age trends and that can reasonably be hypothesised to affect cognate guessing skills: *fluid intelligence*, *crystallised intelligence* and *working memory*. For the sake of completeness, a fourth construct well-known in multilingualism research, viz. *cognitive control*, is briefly discussed as well.

## 3.1 Intelligence

The notion of intelligence is omnipresent in everyday life, but due to its multi-faceted nature, psychologists have found it difficult to come up with a one-size-fits-all definition of what exactly it represents (see Neisser et al., 1996, and Sternberg, 1997). Broadly, the following description may serve as a point of departure for a discussion of the concept:

Intelligence is a very general mental capacity that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—“catching on,” “making sense” of things, or “figuring out” what to do. (Gottfredson, 1997a, p. 13)

Psychometric intelligence, i.e. intelligence as measured by psychological tests, has proven to be an important predictor of educational achievement (Deary et al., 2007; Rohde and Thompson, 2007), professional success (see Gottfredson, 1997b; Neisser et al., 1996; Nisbett et al., 2012) and performance on a gamut of cognitive tasks (see Gray and Thompson, 2004). This raises the question of whether intelligence measures can help to account for inter-individual variance in receptive multilingualism and cognate guessing skills, too.

### 3.1.1 Fluid and crystallised intelligence

A robust observation in intelligence research is that when participants are tested on a battery comprising several intelligence tests, their scores on these tests tend to be positively correlated (see Deary et al., 2006; Neisser et al., 1996). Spearman (1927) took this to suggest that there exists a general factor, termed *g*, that underlies all intelligence test performance. This common factor, extracted using factor analytical techniques, can account for roughly 50 percent of the variance in a wide variety of such tests (Deary, 2001). Despite typically being positively correlated, measures of intelligence also tend to fall into two broad categories with respect to their lifespan trajectories: some measures show a monotonic

age-related decline throughout the better part of adulthood, whereas others remain largely stable or may even show increase.

In general, the intelligence measures that are negatively affected by ageing in adulthood are extracted with reasoning tests in which acquired culture-dependent knowledge plays at best a secondary role. Prime examples of such tests include the Wisconsin card sorting test and Raven's progressive matrices. The factor assumed to underlie a person's performance on largely culture-free reasoning and problem-solving tests is referred to as their *fluid intelligence* (Gf).<sup>6</sup> Measures that are largely robust to age-related decline, by contrast, are typically derived from tests designed to primarily tap the participants' acculturated knowledge, e.g. the Vocabulary, Similarities and Information subsets of the Wechsler Scale. Such tests are taken as indices of the participants' *crystallised intelligence* (Gc). Crystallised intelligence represents the cultural component of intelligence (Kray and Lindenberger, 2007), which is considered to include vocabulary knowledge. Given its culture-dependent nature, Gc is co-determined by education, training and experience.

The Gf–Gc dichotomy is primarily associated with the psychometric theory developed by Cattell and Horn (Cattell, 1963, 1971; Horn, 1982; Horn and Cattell, 1963, 1966), but other theorists have made distinctions along broadly similar lines, e.g. Jones and Conrad (1933, *basic intelligence* vs. *acquired abilities*), Hebb (1942, *intelligence A* vs. *intelligence B*) and Baltes (1987, *mechanics* vs. *pragmatics*). The theoretical assumptions of these different dichotomies do not wholly overlap, but in each case, the facet of intelligence that is liable to age-related decline throughout adulthood is assumed to reflect primarily the biological component of intelligence, whereas the facet that is far more robust with respect to ageing is assumed to reflect primarily the influence of culture. For my present purposes, this is all that matters, and my adopting the Gf–Gc dichotomy is purely a matter of practical convenience that does not carry an endorsement of one theory of intellectual ageing over another.

Further note that the Gf–Gc dichotomy and its associated theory are not without their critics (notably Johnson and colleagues, e.g. Deary et al., 2010; Johnson and Bouchard, 2005; Johnson et al., 2004; Johnson and Gottesman, 2006; Johnson et al., 2008). Their objections mainly con-

---

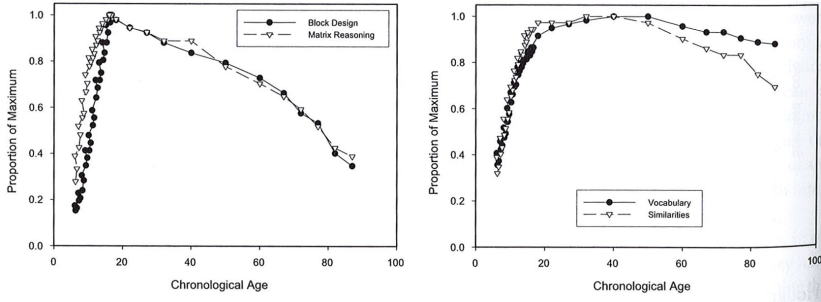
<sup>6</sup>But note that in the light of the so-called Flynn (1987) effect, i.e. the large-scale increases in reasoning task scores over the course of the last decades, the assertion that these tasks measure culture-free reasoning only is untenable (see also Nisbett et al., 2012).

cern the constructs' usefulness in modelling the structure of intelligence and the difficulty of developing tasks that can really tease apart culture-dependent and culture-independent aspects of intelligence (see also Note 6). However, the structure of human intelligence is not a primary concern of this study; what is of interest is the age-patterning of cognate guessing skills. In this context, the observation that reasoning-based intelligence measures and measures of acculturated knowledge display markedly different age trends—which is not questioned by Johnson and colleagues (Deary et al., 2010)—is still highly relevant: the extent to which age-related trends in task performance reflect the developmental trajectories of reasoning-based or acculturated knowledge-based intelligence measures depends on the demands the task in question places on these facets of intelligence—a suggestion put forth by Welford (1958, p. 14, cited in Salthouse, 2006, p. 276) and described by Salthouse (2006) as “probably . . . readily accepted by most contemporary researchers” (p. 276). Thus, the Gf–Gc dichotomy provides useful labels with which the age-patterning of cognate guessing skills can be further broken down into a age-labile component representing reasoning ability and an age-robust component representing acculturated knowledge. It is these developmental trajectories that I will now take a closer look at.

### 3.1.2 Lifespan trajectories

Findings regarding the age trajectory of fluid intelligence are unequivocal: it develops rapidly in childhood and decreases roughly linearly after reaching its peak in early adulthood (see Baltes et al., 1999; Kray and Lindenberger, 2007; Salthouse, 2006). The development of two measures of fluid intelligence is displayed in the left-hand panel of Figure 3.1. It should be noted that age is not the all-determining factor that this graph might suggest it to be: Rabbitt and Anderson (2006) found that age accounts for merely 13.4 to 20.5 percent of the inter-individual variance on three Gf measures in 2,000-participant sample spanning from middle to very old age.

Measures of crystallised intelligence, like those of fluid intelligence, show rapid increase during childhood, but contrary to fluid intelligence, crystallised intelligence stays more or less stable throughout the better part of adulthood or may even show modest improvement (see Baltes et al., 1999; Kray and Lindenberger, 2007; Salthouse, 2006). As regards its development in old age, Singer et al. (2003) found that vocabulary test



**Figure 3.1:** Lifespan trajectories for two measures of fluid intelligence (block design and matrix reasoning; left) and two measures of crystallised intelligence (vocabulary and similarities; right). (source: Salthouse, 2006, Fig. 19.2, p. 276. © Oxford University Press, reprinted with permission. Data from Table A.1 in WASI, 1999)

scores (which serve as proxies of crystallised intelligence), remain largely stable in old age and do not start to decline until age 90, but others (Connor et al., 2004; Salthouse, 2006; Schaie, 1996; Verhaeghen, 2003) report peak vocabulary test performance between the ages of 40 and 60. Gc thus typically increases gradually throughout roughly the first half of adulthood but may decline slightly afterwards. Nonetheless, the contribution of age to inter-individual Gc differences is paltry, ranging from 0.01 to 2.6 percent according to Rabbitt and Anderson’s (2006) data. The overall development of two measures of crystallised intelligence is displayed in the right-hand panel of Figure 3.1.

Note, lastly, that Gc tests aim to measure *general* knowledge, not domain-specific skills and knowledge. The combination of stable Gc and diminishing Gf could suggest that younger adults are, on average, more intelligent than middle-aged and older adults. However, middle-aged and older adults typically have more specialised skill and knowledge resources at their disposal than younger adults, but their advantage in this respect goes unnoticed in Gc tests (Ackerman, 2000).

### 3.1.3 Implications

In the present study, the participants are required to infer the meaning of *Lx* (Swedish) lexical items presented out of context. Lacking context and therefore the possibility to let guide their inferences by contextual information, the participants are forced to derive the meaning of the unknown words on the basis of formal similarity with known words alone. It can be hypothesised that, in order to uncover such similarities and successfully translate the *Lx* items on the basis thereof, the participants will need to muster their ability to deal flexibly with new information (i.e. the *Lx* items) and solve problems creatively, in other words, their fluid intelligence. Additionally, to the extent that cognate guessing relies on abduction (see Section 2.1.2), it involves comparing (orthographic and phonological) patterns and abstracting away from the specifics of such patterns. Participants with high fluid intelligence levels are likely to be more adept at this abstracting process. The likelihood of success on the cognate guessing task is therefore predicted to be positively correlated with the participants' fluid intelligence scores.

Additionally, in order to link unknown word forms in a related language to known words, the participants will also need to draw on their L1 (German) vocabulary knowledge. L1 vocabulary knowledge, generally considered to be part and parcel of crystallised intelligence, can therefore be hypothesised to contribute to success on the cognate guessing task. This ties in with Teleman's (1981) observation that knowledge of relatively infrequent L1 lexemes may be advantageous in inter-Scandinavian communication as these infrequent L1 lexemes sometimes correspond to frequent *Lx* lexemes. By way of example, consider the Norwegian word *begynne* ('to start'), whose meaning cannot be inferred through its typical Swedish translation equivalent *börja*. It can, however, easily be linked to the archaic Swedish form *begynna*.

That said, the L1 words in the Gc test used in this study (see Section 4.2.2 on page 52) are arguably considerably rarer than those that may actually be of use in order to perform well on the cognate guessing task. Insofar as the Gc test provides an adequate proxy of our participants' crystallised intelligence, however, the test scores should also reflect the participants' experience in dealing with L1 material that deviates (lexically, phonetically, phonologically, syntactically, stylistically etc.) from their own L1 norms. Both a rich vocabulary and experience in dealing with deviating L1 material are likely to be useful when trying

to make sense of words in a related but unknown language (see Teleman, 1981). This prediction is consistent with findings that cognate guessing skills are higher in participants who have a dialect background or who know multiple languages that are closely related to one another as well as to the unknown language (see Section 2.1). Moreover, a large linguistic repertoire provides greater potential for forming associations (Kyllonen et al., 1991), on the basis of which more accurate speculations about likely sound or grapheme correspondence rules can be generated by means of abduction. Crystallised intelligence, and L1 vocabulary knowledge more specifically, is therefore likely positively associated with performance on cognate guessing tasks.

To conclude, I reiterate that the overall age-patterning of our participants' cognate guessing skills is likely to depend on the cognate guessing task's demands on fluid and crystallised intelligence, respectively (as per Welford, 1958). If cognate guessing is largely dependent on crystallised resources, it should be largely stable throughout the adult lifespan; if cognate guessing is largely dependent on fluid resources, it should show an age-related decline in adulthood. In either case, an age-related increase throughout childhood and adolescence is to be expected, which is what previous studies have indeed shown (see Section 2.4).

## 3.2 Working memory

A central construct in cognitive psychology, working memory (WM) refers to the ability to briefly maintain and manipulate new information (Baddeley and Hitch, 1974; Hitch, 2006; Park and Payer, 2006, among many others). It should not be confused with short-term memory (STM): whereas STM is conceived of as a mere information storage device, WM comprises both storage and information processing capabilities. Both the STM and WM constructs contrast with long-term memory (LTM), which entails the more permanent storage of information. Several theoretical models of WM have been developed throughout the years, and I will base my discussion on Baddeley and Hitch's (1974) highly influential multi-component WM model.

### 3.2.1 The Baddeley–Hitch multi-component model

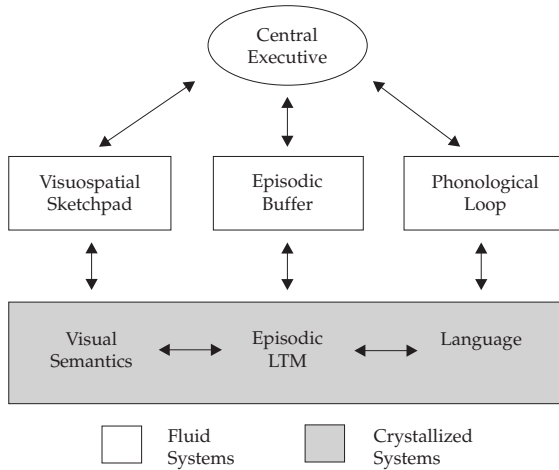
Whereas other scholars, most notably Atkinson and Shiffrin (1968), had earlier conceptualised WM as a unitary device capable of both short-term storage and manipulation of information (see also Baddeley, 2003a), Baddeley and Hitch (1974) proposed a three-component WM model. This made it possible to explain neuropsychological and experimental data indicating that deficiencies or disruptions affecting one aspect of WM (e.g. short-term storage) need not necessarily result in a dramatic break-down of another WM aspect (e.g. manipulation of information). If WM functioned as a unitary system, such differentially affected aspects would be much harder to explain (see Baddeley, 2003a). The multi-component model has been refined through the years and now consists of four rather than the original three components.

Figure 3.2 on the next page graphically presents the current iteration of the multi-component WM model. It features three so-called ‘slave systems’, viz. the *phonological loop*, the *visuospatial sketchpad* and the *episodic buffer*, as well as one control system, the *central executive*. The phonological loop (originally called *articulatory loop*) is a subsystem dealing with verbal information.<sup>7</sup> It consists of a passive phonological store, which is capable of storing verbal information for the duration of about two seconds, and an active (subvocal) articulatory rehearsal component (see Baddeley, 2000; Baddeley and Hitch, 2000; Baddeley et al., 1984; Gathercole and Baddeley, 1993; Repovš and Baddeley, 2006). The latter component allows verbal information in the phonological store to be refreshed, thereby lengthening its longevity in the store. The phonological loop plays a key role in L1 vocabulary acquisition (see Gathercole and Baddeley, 1993) as well as foreign language vocabulary learning in youngsters (Cheung, 1996; Service, 1992; Service and Kohonen, 1995; all cited in Baddeley et al., 1998) and adults (Martin and Ellis, 2012; Papagno et al., 1991; Papagno and Vallar, 1992). Moreover, it may be involved in grammar acquisition in the L1 (Adams and Gathercole, 1995, 1996; Blake et al., 1994; Daneman and Case, 1981) and in the L2 (for references, see Martin and Ellis, 2012, pp. 381–382).

The second slave system, the visuospatial sketchpad (sometimes *scratch pad*), is “assumed to be capable of temporarily maintaining and

---

<sup>7</sup>In this discussion, digit names are considered to be snippets of verbal information. As such, ‘verbal’ contrasts with ‘visuospatial’. Note, however, that Daneman and Merikle (1996), among others, contrast verbal WM tasks with math-based WM tasks.



**Figure 3.2:** The current multi-component working memory model featuring the episodic buffer. (source: Baddeley, 2002, Fig. 3, p. 93. © American Psychological Association, reprinted with permission.)

manipulating visuospatial information” (Baddeley, 2002, p. 88). The sketchpad’s involvement in language processing and production is limited (Baddeley, 2003a; Gathercole and Baddeley, 1993) and I will therefore not discuss this component of the Baddeley–Hitch WM model any further.

The third slave system, the episodic buffer, is the latest addition to the multi-component model (Baddeley, 2000). This new component is “assumed to be a limited-capacity temporary storage system that is capable of integrating information from a variety of sources” (Baddeley, 2000, p. 421), including the other two slave systems and long-term memory. These snippets of information from different sources are assumed to be bound together into “coherent episodes” (Baddeley, 2000, p. 421), which are then retrievable through conscious awareness. Like the other slave systems, it can not only incorporate stored (crystallised) representations in long-term memory, but information temporarily stored and manipulated in it can be updated or stored more permanently in long-term memory, hence the double arrows in Figure 3.2. Since the episodic buffer is a relatively new addition to the multi-component WM model, research on it is “still in its infancy” (Repovš and Baddeley, 2006, p. 16). Nonethe-

less, given its role in retrieving information from long-term memory, it can be assumed that the buffer is crucial in lexical retrieval during speech recognition. Rudner and Rönnerberg (2008) consequently suggested that the functioning of the buffer may become particularly strained if the lexical items stored in LTM and the phonological information in the speech signal do not match, as is for instance the case in noisy conditions.

The three slave systems are controlled by the central executive, which is “assumed to be responsible for the attentional control of working memory” (Baddeley, 2003a, p. 201). More specifically, it is assumed to focus the limited attentional resources to relevant new information as well as to divide these resources among the slave systems, for instance when attending to two separate tasks (Baddeley, 2002). In the current iteration of the Baddeley–Hitch model, the central executive does not itself have a storage capacity (see Baddeley and Logie, 1999). While considered the cornerstone of the multi-component model, the central executive is admittedly the least understood of all components (Baddeley, 2003b). Executive processing may contribute somewhat to language comprehension (Daneman and Merikle, 1996), but all in all, the role of the central executive (in its current specification) in language processing appears to be a comparatively under-researched topic.<sup>8</sup>

To conclude this discussion, I point out that while the multi-component model of working memory is arguably the most influential account of working memory and short-term recall, criticisms have been levied against the specification of its individual components, for instance the episodic buffer (e.g. Ruchkin et al., 2003) and the phonological loop (e.g. Campoy, 2008; Jones et al., 2006, 2007, 2004; Romani et al., 2005), and other WM models obviously do exist (e.g. Just and Carpenter, 1992). That said, the purpose of the present study is not to argue in favour of any one such theory. Rather, its primary aim is to investigate how the ability to infer the meaning of unknown foreign-language words develops throughout the lifespan. Insofar as WM performance varies as

---

<sup>8</sup>This statement warrants an aside. Martin and Ellis (2012, pp. 382–383), for instance, cite several studies investigating the link between L2 learning and working memory, and it is not this relationship as such that is relatively under-researched. But as far as I can gather, the authors cited by Martin and Ellis were not concerned with singling out the contribution of executive processing to L2 learning, but rather the impact of WM as a whole. Interesting though these findings are, they cannot directly be interpreted as pertaining to executive processing specifically (see also Section 3.2.2).

a function of age and to the extent that WM can be linked to language processing and learning, WM theory can contribute to our understanding of the development of cognate guessing skills throughout the lifespan. Criticisms notwithstanding, the multi-component model provides an elegant framework within which the possible involvement of WM in cognate guessing can be discussed.

### 3.2.2 Measuring working memory capacity

Typically, so-called working memory span tasks (or complex span tasks) are used to determine the participants' working memory capacity. A prime example of a WM span task is the reading span task developed by Daneman and Carpenter (1980), which requires subjects to first read out loud a number of visually presented sentences and then recall the final words of these sentences. One of many alternatives and variations is the operation span task (Turner and Engle, 1989), which requires subjects to check whether a simple mathematical equation is correct whilst remembering the last numbers in these equations. The reading span and operation span tasks are verbal in nature, as are alternatives such as the backward digit span task, in which subjects have to store a series of digits and repeat them back in inverse order. Visuospatial WM span tasks have, however, also been developed (e.g. Shah and Miyake, 1996).

All WM span tasks require subjects to simultaneously process ("transform or manipulate", Park and Payer, 2006, p. 129) and store new information, and WM span measures have been taken to reflect the quality of the central executive in the framework of the multi-component WM model (Gathercole and Baddeley, 1993, pp. 72–73). It is worth pointing out, however, that these span tasks were not developed to tap the central executive exclusively. Daneman and Carpenter (1980), who developed the first span task, did not adhere to the Baddeley–Hitch multi-component WM model. Mapping their conceptualisation of working memory capacity onto a specific component in the Baddeley–Hitch model is therefore unsound. Since the respecification of the central executive as a component without storage capacity by Baddeley and Logie (1999), verbal WM span tasks may perhaps better be thought of as composite tests measuring partly executive processing efficiency, partly the quality of the phonological loop and partly additional aspects

of cognition, depending on the specifics of the task (see Bayliss et al., 2005).<sup>9</sup>

### 3.2.3 Lifespan trajectories

Park and Payer (2006) tracked the developmental trajectories of overall WM and STM performance throughout the lifespan. They found that while both WM and STM measures show regular decline throughout the adult lifespan, WM measures are more age-sensitive than STM measures, i.e. WM shows steeper decline. In combination with rapid increases in both WM and STM in childhood (see Gathercole, 1999), these developmental trajectories bear a striking resemblance to the age-patterning of fluid intelligence capabilities (see Section 3.1.2). This similarity is not merely superficial, as evidenced by strong correlations between WM and Gf measures (Ackerman et al., 2002; Colom et al., 2004; Conway et al., 2002; Engle et al., 1999; Kane et al., 2005; Kyllonen and Christal, 1990; Oberauer et al., 2005; Süß et al., 2002; see also Heitz et al., 2006), which may be attributable to a common reliance on attentional control (see Section 3.2.4) and to the impact of processing speed on both constructs (see Salthouse, 1996, 2000; Salthouse and Babcock, 1991; Salthouse and Meinz, 1995; Verhaeghen and Salthouse, 1997).

From a multi-component WM model point of view, the rise and fall of overall WM performance raises the question of which individual components are implicated in this development. The answer seems to be that the phonological loop (Gathercole and Baddeley, 1993; Hitch, 2006; Salthouse, 1994), the visuospatial sketchpad (Riggs et al., 2006; Salthouse, 1994), the central executive (Gathercole, 1999; Park and Payer, 2006; Salthouse, 1994; Salthouse and Babcock, 1991) and the episodic buffer (Kessels et al., 2007; Smith, 2005) all seem to be involved to some extent.

---

<sup>9</sup>My hedging in stating what precisely it is that a complex span task taxes is justified judging by Hitch's (2006) observation that while "a useful tool for studying individual differences in working memory", "[it is not as yet] a thoroughly well understood task" (p. 115).

### 3.2.4 Implications

For my present purposes, the age-related WM development is particularly relevant considering its importance in accounting for inter-individual differences on higher-order cognitive tasks. The role of WM in language learning and processing has already been touched upon in Section 3.2.1. In addition, WM task performance has been found to play a role in a host of higher-order cognitive tasks as diverse as complex learning, mental arithmetic and reasoning (see Engle et al., 1999). Here, I propose that the participants' performance on a WM task may similarly be positively correlated with their performance on the cognate guessing task and that the overall age-patterning of cognate guessing skills may therefore partly be influenced by the development of WM.

Why is WM involved in such a wide array of tasks and why do I expect it to be involved in cognate guessing, too? Engle and colleagues (e.g. Conway et al., 2003; Engle, 2002; Engle et al., 1999; Heitz et al., 2006; Kane et al., 2001) have argued that the centrality of working memory in a wide variety of tasks is primarily due to its role in focussing one's attention on what is relevant in a given situation. Kane et al.'s (2007) finding that high WM individuals are more likely to stay focussed during a difficult task corroborates this view. In terms of the multi-component model, the 'focussed attention' account of individual differences in task performance attaches the highest importance to the central executive. However, I suggest that any link found between WM and cognate guessing need not solely be due to a shared reliance of both tasks on the ability to focus one's attention. Rather, the quality of the phonological loop may be a major determining factor. To understand why the phonological loop may play a crucial role in cognate guessing, consider what processes the participants need to engage in when performing this task in the two modalities (auditory and visual).

Broadly speaking, the cognate guessing task can be assumed to require the participants to retrieve several known vocabulary items (possibly from different language varieties) from LTM as potentially useful transfer bases via which the  $Lx$  stimuli's meaning may be inferred. The search for the most suitable transfer bases may then entail mentally transforming the orthography or phonology of several potential transfer bases and of the  $Lx$  stimulus at hand and comparing the results. In the context of rhyme judgement experiments, Besner (1987) suggests that similar "post-assembly phonemic segmentation and deletion processes"

(p. 474) tax the phonological loop. Thus, tasks requiring subjects to carry out segmental operations on two phonological forms whilst maintaining them in the phonological loop are likely to place great demands on the phonological loop. Participants with a large phonological loop capacity may therefore be able to retrieve, transform and compare more forms at the same time without losing track of them, which may result in their making more informed guesses. Moreover, in the case of auditory stimuli, a well-functioning phonological loop may help to preserve the stimuli themselves during the retrieval, transformation and comparison processing. In the case of visually presented stimuli, this factor is presumably of less importance if they are displayed throughout.

Apart from the central executive and the phonological loop, the episodic buffer may also be involved in cognate guessing. Doetjes and Gooskens (2009) and Schüppert et al. (2010, see Schüppert, 2011, Ch. 6) present correlational and neurological evidence, respectively, that suggests that participants in auditory cognate guessing tasks do not only rely on phonological representations but also partly on their orthographic knowledge. Speculatively, auditory cognate guessing entails the transformation and comparison of integrated phonological and visual-orthographic representations, which would engage the episodic buffer. Equally speculatively, visual cognate guessing tasks might also encourage the comparison of such integrated representations. Indeed, Möller and Zeevaert's (2010) finding that participants overtly or subvocally generate their own phonological representations of the items to be inferred (which I can confirm on the basis of my own data collecting experience) is not necessarily at odds with the hypothesis that they nevertheless partly rely on concurrently maintained visual-orthographic representations as well.

Admittedly, the present study will not be able to address these speculations regarding the locus of any expected association between WM performance and cognate guessing skills. But if such an association does indeed exist, these speculations may be followed up in future experimental studies.

### 3.3 Cognitive control

A fourth cognitive variable (in addition to fluid and crystallised intelligence and working memory) that I considered in my investigation was

*cognitive control*, i.e. the ability to direct attention to task-relevant information and away from task-irrelevant information.<sup>10</sup> Cognitive control is often measured by means of the Stroop (1935), Simon (1969) and flanker (Eriksen and Eriksen, 1974) tasks and is typically considered to be semi-synonymous to *attentional*, *executive* or *inhibitory control*, which are in principle different (Braver and Barch, 2002) though highly related (Smith and Jonides, 1999) constructs.

Cognitive control is a well-known dependent variable in multilingualism research thanks to a host of studies investigating the cognitive effects of bi- and multilingualism, notably by Bialystok and colleagues (e.g. Bialystok et al., 2005a, 2004, 2008, 2005b; Emmorey et al., 2008; Luk et al., 2011; Martin-Rhee and Bialystok, 2008; see also, e.g., Costa et al., 2009; Hilchey and Klein, 2011; Morton and Harper, 2007, 2009). Furthermore, the ability to efficiently focus one's attention is often considered central to higher-order cognitive performance (e.g. Conway et al., 2002, 2003; Engle, 2002; Engle et al., 1999; Heitz et al., 2006; Kane et al., 2007, 2005), and cognitive control is known to show age-related increase and decline (Bialystok et al., 2004, 2005b; Bugg et al., 2007; Schiller, 1966; Utzl and Graf, 1997; Van der Elst et al., 2006; Van der Lubbe and Verleger, 2002; West, 1999). For these three reasons, cognitive control seemed an interesting variable to include in the present investigation, and a Simon task was added to the task battery (see Section 4.2.7).

To anticipate the results, however, the Simon task yielded a measure of poor quality inasmuch as nearly half of the participants did not show the critical effect (see Appendix B). Moreover, Paap and Greenberg (2013) have recently shown that cognitive control measures extracted using one kind of task (e.g. the Simon task) do not predict cognitive control measures extracted using another kind of task (e.g. flanker task), i.e. they have poor external validity. For these reasons, cognitive control was not further considered in the analyses and will not further be discussed here.

---

<sup>10</sup>Note that Berthele (2008, 2011) proposed that a key skill in cognate guessing is knowing when to quit looking for increasingly less suitable transfer bases once one or a couple candidates have been identified. Although Berthele called this ability *Fokussierungsfähigkeit*, no conceptual overlap between this ability and attentional focussing or cognitive control is implied.



# Chapter 4

## Method

In the previous chapters, I discussed several linguistic and cognitive factors that may shape the lifespan development of cognate guessing skills. In what follows, I turn my attention to an empirical study designed to track this lifespan development and identify its linguistic and cognitive driving forces. This chapter discusses the design of this investigation, Chapters 5 to 7 present its results with respect to inter-individual differences in cognate guessing skills, and Chapter 8 discusses the implications of these results. Note that the investigation of the impact of item-related characteristics on cognate guessing and their interaction with participant-related variables is the subject matter of Part III.

### 4.1 Participants

The three psycholinguistically oriented subprojects of the *Multilingualism through the lifespan* project (see Section 1.4.1) set out to recruit participants across the adult lifespan as well as children aged approximately 11 years and adolescents aged approximately 15 years old. To be eligible for participation, candidates had to consider themselves to be native speakers of a Swiss German dialect. Candidates who considered themselves to be native speakers of one or more languages besides Swiss German could also participate in the study. Since the cognate guessing task featured Swedish words, candidates with self-reported knowledge

**Table 4.1:** Main demographic characteristics of the participant sample.

Age group	<i>n</i> total	<i>n</i> women	<i>n</i> men	Mean age
10–12	23	9	14	10.6
14–16	19	11	8	15.4
20–29	20	12	8	26.0
30–39	20	14	6	33.6
40–49	21	18	3	43.7
50–59	19	11	8	55.0
60–69	19	11	8	64.6
70–79	21	4	17	72.3
80+	5	1	4	82.6
Overall	167	91	76	41.0

of Swedish or any of the related North Germanic languages (Danish, Faroese, Icelandic, Norwegian) were filtered out a priori. Language experts such as language or linguistics students or interpreters were likewise filtered out a priori as their meta-linguistic knowledge, especially about historical sound laws, could have skewed the result severely.

We managed to recruit 167 participants in total, all of whom reported normal or corrected-to-normal vision and hearing. Participants were financially compensated for their participation and their travelling expenses and gave their written informed consent before participation. The participants' main characteristics are given in Table 4.1. As can be gleaned from this table, participants aged 80 or older turned out to be difficult to recruit. Moreover, the age groups 10–12, 70–79 and 80+ are dominated by men, whereas the other age groups show a majority of women. When investigating age trends in this sample, it is therefore necessary to reckon with the participants' sex as a potential confound variable.

## 4.2 Tasks and procedure

All 167 participants were recruited in order to provide data for three subprojects. While this brings substantial advantages in terms of participant recruitment, data collection and comparability between the

subprojects, these advantages come at a certain price: out of deference to the youngest and the oldest participants in particular, the entire task battery could not be tiresomely long and the time available had to be shared among three subprojects. Consequently, the tasks had to be fairly short and there was no time to carry out several tests tapping the same cognitive construct more elaborately for a latent variable analysis. Furthermore, several independent variables were extracted that are not of primary interest to the project discussed in this thesis. Some other independent variables turned out to be of dubious quality and were not considered when modelling the lifespan development of cognate guessing skills. For the sake of completeness and scientific propriety (see Simmons et al., 2011), even the independent variables that were not used for the analyses will nevertheless briefly be presented.

Data collection sessions lasted approximately two and a half to three hours and took place in a quiet room at the participants' convenience. The participants were tested individually or in groups of no more than five. Table 4.2 on the following page presents the order of the tasks administered. Note that Tasks A, B, C and D served to collect data for specific subprojects. These four tasks were presented in varying order so as to offset fatigue effects.

### **4.2.1 Language background questionnaire**

Before proceeding with the task battery, the participants completed a language background questionnaire. They self-assessed their reading and listening skills in standard German, English, French and any other language they knew using the self-assessment grid for the Common European Framework of Reference for Languages (CEFR; available from <http://europass.cedefop.europa.eu/de/resources/european-language-levels-cefr>). Possible levels range from 'no knowledge' over 'A1' or beginner to 'C2' or mastery.

Six further variables were extracted using the questionnaire that were not of primary interest for my purposes: (1) the participants' overall interest in language(s); (2) their manner of acquisition or learning of the languages they knew; (3) their self-estimated aptitude for learning new languages; (4) which aspect of language learning they found hardest and easiest: grammar, vocabulary or pronunciation; (5) the frequency with which they used the languages in their repertoire; and (6) how much

**Table 4.2:** Sequence of tasks in the task battery. Tasks A, B, C and D refer to tasks specific to the three subprojects and were administered in varying order in order to offset fatigue effects.

German vocabulary test (WST)
German lexical decision task
Task A
Task B
Pause (15')
Task C
Task D
Pause (10–15')
English test
Simon task
Backward digit span task
Raven's advanced progressive matrices

they liked to use foreign languages in general. These variables were not taken into consideration in this thesis and will not further be discussed.

### 4.2.2 German vocabulary test

A measure of crystallised intelligence was extracted using a German-language vocabulary test (WST; Schmidt and Metzler, 1992). This test, which was administered as a paper-and-pencil task, consists of 42 series of words and non-words. The participants' task is to tick the existing German word presented alongside five orthographically and phonotactically permissible non-words. The target words ranged from the educated but common, e.g. *Ironie* 'irony', to the highly arcane, e.g. *Heddur*, an aluminium alloy. The participants were explicitly instructed not to guess. One point was awarded for each correctly identified target word.

### 4.2.3 English language proficiency test

The participants' English language proficiency was assessed by means of a 20-item multiple choice grammar test (the first 20 items from the Oxford Placement Test; Allen, 1992) and a 25-item C-test (*Increasing your confidence in listening*, available from the website of the Language Centre of the University of Rostock: <http://www.sprachenzentrum.uni-rostock.de/einstufungstests/c-test/c-test-englisch/>, last accessed 8 March 2013). The reason why grammar rather than vocabulary tests were used was that another subproject focussed on pragmatic strategies for coping with grammatically ambiguous English sentences. That said, L2 grammar test scores and L2 vocabulary test scores are usually substantially correlated (see e.g. Shiotsu and Weir, 2007), and these tests should thus provide an adequate proxy of English vocabulary knowledge, too.

### 4.2.4 Backward digit span task

Working memory capacity was measured using a German-language backward digit span task (BW-DS; Tewes, 1991, pp. 53–54). In a BW-DS, the participants are presented with spoken digit sequences that they need to repeat back verbally in reversed order. The length of the sequences increases from two to eight digits, with two sequences for each level. The participants proceed to the next level if they produce at least one wholly correct backward repetition at a given level. The task is aborted if they fail to provide at least one out of two backward sequences at a given level, or after the second eight-digit sequence, whichever comes first. The BW-DS results in two measures: a 'span' measure ranging from 2 to 8 indicating the highest level at which at least one correct answer was provided and a 'total' measure ranging from 1 to 14 indicating the total number of correct responses.

The BW-DS is quick and easy to administer and commonly used to measure working memory capacity (e.g. Daneman and Merikle, 1996; Gathercole et al., 2004; Grüter and Crago, 2012; Kormos and Sáfár, 2008). The use of the BW-DS as a task for measuring WM capacity is motivated on the grounds that participants have to both store and manipulate information. Some authors (e.g. Engle et al., 1999, and Park and Payer, 2006) have nonetheless argued that it is a measure of short-term memory capacity instead, given that it groups with typical short-term memory

tests in factor analyses. However, a pilot run with an arguably more valid WM task, the automated operation span task (Unsworth et al., 2005), indicated that the latter task would be prohibitively gruelling mentally and time-wise, particularly for the youngest participants, and we therefore settled on the BW-DS.

In order to keep the digit sequence presentation rates constant and to avoid general experimenter effects (caused by, for instance, my own foreign accent), the digit sequences were prerecorded in standard German by a female native speaker of Swiss German<sup>11</sup> and were presented to the participants over headphones.

### 4.2.5 Raven's advanced progressive matrices

The participants' fluid intelligence was measured using the second set of Raven's advanced progressive matrices (Raven, 1962). This set contains 36 abstract puzzles in which eight patterns are presented in a 3-by-3 grid. The task is to select the missing ninth pattern that fits logically within the sequence from a list of eight possible patterns presented underneath the grid. The participants were not explicitly discouraged from guessing. One point was awarded for each correctly selected pattern.

### 4.2.6 Cognate guessing task

#### Materials

The participants were presented with a computer-run cognate guessing task. This task consisted of two blocks featuring 50 Swedish words each. One block was presented visually and the other block aurally. As the goal was to establish how well the participants were able to understand words in an unknown foreign language that are genealogically related to words in languages they do know (specifically German, English and French),  $2 \times 45$  Swedish words with German, English or French translation-equivalent cognates were selected from the Swedish vocabulary. These words differ greatly both in the corpus frequency of their German, English or French cognates and in their degree of formal overlap with their cognates (as will be discussed in Chapter 9 in Part III). These words are referred to as

---

<sup>11</sup>I thank Barbara Ruf of the Department of Multilingualism, University of Fribourg, for recording these sequences.

‘target stimuli’ and are presented in Tables A.1 and A.2 in Appendix A on page 189.

It bears pointing out that the target stimuli were not selected by means of any independently replicable sampling procedure. Instead, their selection was guided by prior experience with cognate guessing tasks, fine-tuned by a pilot run with 19 Swiss-German students. This allowed me to select stimuli that would be translatable by some varying proportion of the participant sample whilst avoiding floor and ceiling effects. In concrete terms, the number of words showing complete formal overlap with their German, English or French cognates was limited in order to avoid ceiling effects. Similarly, the number of relatively short spoken stimuli featuring fricativised onsets was limited so as to avoid floor effects, as these were found to be well-nigh incomprehensible to the pilot participants.<sup>12</sup>

Five stimuli per block did not have any German, English or French translation-equivalent cognates. These ‘profile stimuli’, to borrow a term from the EuroCom vocabulary (Klein and Stegmann, 2000), were included in order to allow me to verify whether the participants did indeed not have any prior lexical knowledge of Swedish: since their meaning cannot be inferred through their formal resemblance to one or more words in a known language, participants who are able to translate them correctly can be assumed to have some prior, albeit limited, lexical competences in Swedish (for a similar use of non-cognates in a cognate guessing task, see Kürschner et al., 2008). The profile stimuli, all highly frequent in Swedish, are also presented in Tables A.1 and A.2.

## Procedure

The Swedish cognate guessing task was administered with E-Prime 2.0 (Schneider et al., 2002). The 50 spoken stimuli were recorded by a native speaker of (Central) Standard Swedish and were presented over headphones.<sup>13</sup>

The participants were told that they would be presented a series of Swedish words, some but not all of which they might be able to

---

<sup>12</sup>Swedish /k/ and /sk/, for instance, are typically realised as [k] and [sk], respectively, when followed by a front vowel. Examples include *kämpa* [ˈjɛmpa] ‘to fight’ and *skinn* [ʃɪn] ‘skin’.

<sup>13</sup>I thank Kristina Borgström, Psychology Department, University of Lund, for recording these items.

understand. They were instructed to strike the ‘J’ key on the laptop’s keyboard if they thought that they could understand the word presented and the ‘F’ key if they thought that they could not (vice versa for left-handed participants). If they indicated that they might understand the stimulus, a text prompt would appear in which they could enter a translation suggestion in German.

The two blocks (written and spoken) were presented in randomised order and within each block, the items were presented in randomised order as well. Each trial consisted of a 1,000 ms fixation phase during which a fixation cross (‘+’) was presented in the centre of the screen (Courier New, 18 pt). After the fixation phase, written stimuli were presented in Courier New (18 pt) in the centre of the screen and spoken stimuli were played once through both channels of the headphones. Visual stimuli remained on the screen until the participant struck the ‘F’ or ‘J’ key. Response latencies were measured from stimulus onset onwards until the participant struck the ‘F’ or ‘J’ key. These response latencies were used only for the purposes of filtering out data points with unrealistically fast latencies that could likely be attributed to inadvertent keystrokes. A ‘J’ stroke prompted a text box in which a translation suggestion could be entered. The text box remained on the screen until the participant confirmed the translation suggested by pressing ENTER. Intertrial intervals lasted 1,500 ms.

Before each block, a training run with 5 stimuli took place. After this training run, participants could notify the experimenters in case it emerged that they had not fully understood the instructions or in case of technical difficulties (e.g. low volume). Participants did not have to score perfectly in order to proceed to the actual task and were not informed about the correctness of their translations.

## Scoring

All translations were checked and coded binarily for their correctness; I did not judge the reasonableness of the answers. Indeed, a great deal of answers were completely reasonable but just happened to be incorrect translations of the Swedish stimulus (e.g. *Kniff* ‘pinch’ for *kniv* ‘knife’ or *beerdigen* ‘to bury’ for [‘bœrja’ ‘to begin’]), whereas others were not too far off to begin with and would in all likelihood have been correct had some context been provided (e.g. *sprechen* ‘to talk’ for *språk* ‘language’).

When checking the translations, I entirely disregarded capitalisation. I did not consider misspelt words to be wrong translations as long as the misspelling did not give rise to another existing word. Thus, *keiser* was accepted as a correct translation for the written stimulus *kejsar* ‘emperor’, even though the correct orthography is *Kaiser*. The translation *Grippe* for the spoken stimulus [grøp] (*grupp* ‘group’), on the other hand, was not considered an acceptable translation: even though it may well have been a misspelling of the correct German translation *Gruppe* (‘u’ and ‘i’ lie right next to each other on the keyboard), *Grippe* is an existing word in German, meaning ‘influenza’. In contrast to Kürschner et al. (2008), who defined spelling errors as “instances where only one letter had been spelt wrongly without resulting in another existing word” (p. 85), I did not apply a rigorous definition of spelling errors but considered each case individually.

In case the translation provided did not perfectly match the model translation or a synonym, I operated along the following lines:

- If more than one translation was provided, the answer was rated as correct if one of the translations was correct. For instance, the answer *denken oder trinken* ‘to think or to drink’ for [ˈtɛŋkə] ‘to think’ was rated as correct.
- Even though all nouns were presented in the singular, both (nominative) singular and plural translations were accepted, e.g. *Blumen* ‘flowers’ for [ˈblʊmə] ‘flower’.
- Even though all verbs were presented in the infinitive, I accepted translations in the infinitive, imperative and simple present. Thus, *sitz* ‘sit (imp.)’, but also: ‘seat’ was considered a correct translation of *sitta* ‘to sit’.
- Even though all adjectives were presented in their predicative forms, attributive forms were also accepted, e.g. *erste* ‘first (attr.)’ for [føɛst] ‘first (pred.)’.
- French and English translations were accepted as well.

I was, however, less forgiving as far as ‘near-miss’ translations were concerned. The translation *Zirkel* ‘circle’ for *cyckel* ‘(bi)cycle’ was therefore rated as incorrect. Likewise, hyper- and hyponyms of the correct translation were rated as incorrect: neither *Kraftwerk* ‘power

station’ for [ˈfæ:ŋkraʃt,ʋɛrk] ‘nuclear power station’ nor *taschenmesser* ‘pocket knife’ for *kniv* ‘knife’ were accepted as correct answers. Moreover, only translations belonging to the same part of speech as the model translation were accepted, i.e. the noun *Rhythmus* ‘rhythm’ was not considered an acceptable translation of the adjective *rytmisk* ‘rhythmic’. Exceptions to this rule were cases in which, for instance, the imperative of the correct translation of a verb stimulus was identical to a related noun (e.g. *sitz* ‘sit (imp.), but also: seat’ for *sitta* ‘to sit’) or the nominative plural of the correct noun was identical to a verb form (e.g. *Schminken* ‘make-up (pl.), but also: to apply make-up’ for [smɪŋk] ‘make-up (sg.)’), as per the rules outlined above.

#### 4.2.7 Measures not used in the analyses

The task battery featured two tasks that extracted measures which I did not use in the analyses. First, some of the other subprojects’ research questions concerned age trends in response latencies on particular tasks. Response latencies on verbal tasks show age trends even in the L1, and increases in average response latencies on foreign-language verbal tasks beyond middle age may potentially be attributed to a parallel L1 trend. A German-language lexical decision task was therefore included in the design in order to extract an L1 latency control measure. This measure was not used in the analyses in this thesis, however, and will not further be discussed.

Second, in order to extract a measure of their cognitive control for reasons discussed in Section 3.3, a Simon task was included in the task battery. The Simon task is an experimental design in which participants are required to make a spatial response, i.e. a left or right key-press, cued by a non-spatial stimulus characteristic (e.g. stimulus colour) but irrespective of the stimulus’ location of presentation (Simon and Small, 1969). Even though the location of presentation is irrelevant to the required response, participants tend to react faster when the response location is congruent with the location of presentation than when it is not. This effect is known as the *Simon effect*.

The Simon task enjoys some popularity in the psycholinguistic literature on bi- and multilingualism as a means of extracting a measure of cognitive control (see Section 3.3). In these studies, the size of the Simon effect is assumed to reflect the participants’ ability to focus on the task goals by ignoring irrelevant information: participants with smaller Simon

effects are better able to do this than are participants with larger Simon effects. The Simon task used in the present study failed to produce a reliable Simon effect, however, and was not further considered when modelling the inter-individual differences in cognate guessing skills. For more details on the task design and its results, refer to Appendix B on page 195.

## 4.3 Method of analysis: Mixed-effects modelling

The analyses presented in this thesis rely heavily on statistical tools called the *generalised linear mixed-effects model* (GLMM) and the *generalised additive mixed-effects model* (GAMM). GLMMs have come to the fore in the social sciences in recent years, but not all readers may be familiar with them. In what follows, I will first provide a non-technical but fairly extensive account of mixed-effects modelling using GLMMs. Second, since GAMMs are even more of a novelty in the social sciences, I will briefly explain what GAMMs do and how their output can be interpreted. For this introduction, I assume familiarity with traditional analytical tools such as ordinary (least squares) and logistic regression modelling, accessible introductions to which are provided by Baayen (2008) and Johnson (2008).

### 4.3.1 A gentle introduction to the generalised linear mixed model

Consider a hypothetical receptive multilingualism study similar to mine in which 50 participants are presented with 30 foreign-language ( $Lx$ ) stimuli that they have to translate into their L1. The researchers may want to find out to what degree the number of languages that the participants know affects their translation performance. When using traditional methods, such as correlation and regression analyses, they would most likely calculate the total number (or the proportion) of correct translations per participant and use this tally as their dependent variable. If, on the other hand, the researchers would like to find out to what degree item-related factors such as stimulus length affect the stimuli's intelligibility, they would presumably calculate the total number

(or proportion) of correct translations per stimulus to use it as their dependent variable.

On the face of it, both analyses seem entirely defensible. However, when performing inferential statistical analyses, what is of interest is knowing whether the trends observed in our data sample stand a good chance of generalising to a wider population. If we were to find a statistically significant effect of stimulus length on stimulus intelligibility, we would believe by implication that the length effect should also hold for stimuli other than the 30 items used in the task. Likewise, if we were to find a significant effect of linguistic repertoire size on translation performance, we would ipso facto draw the inference that this effect matters not only to our 50 participants but more generally to the population from which they were sampled.

But there is a problem: if we found a significant effect of stimulus length across the 30 items, then all that this would technically mean is that we could for now assume that stimulus length generally affects intelligibility *in the 50 participants*. Likewise, if we found a significant effect of linguistic repertoire size across our 50 participants, we could only draw the inference that it generally affects translation performance *for the very same 30 stimuli used in the task*. The reason is that, as Coleman (1964) and Clark (1973) pointed out, both the participants and the stimuli were sampled from larger populations, yet we did not acknowledge this in our analyses. While the hypothetical length effect may generalise to a wider population of stimuli, we have strictly speaking no evidence that it may generalise to a wider population of participants, and vice versa for the linguistic repertoire effect. It is, of course, quite possible that these effects would indeed generalise, but technically we simply do not know. More subtly, the effects may be present in those wider populations, but their strength may well differ considerably from stimulus to stimulus or from participant to participant.

Using traditional correlational and regression techniques, we cannot straightforwardly escape what Clark (1973) dubbed the *language-as-fixed-effect fallacy*: if we do not aggregate the data over participants or stimuli but run a logistic regression on the raw data instead, we blatantly violate the model's assumption that the data points (or, more accurately, the model's errors) be independent of one another. For ANOVA-based analyses, a few statistical methods have been developed in order to determine whether observed trends in psycholinguistic data are indeed

likely to hold across both participants and items, but these approaches are not optimally suited when dealing with continuous predictor variables or non-continuous outcome variables (Jaeger, 2008).

A more suitable alternative are (generalised) linear mixed-effects models, which have come to the fore as tools for psycholinguistic analyses in recent years. Linear mixed-effects models have a number of advantages over the more traditional ANOVA-based approaches. First, they can cope with binary outcome variables by using the logistic function as a link function, thus eliminating the need to average and transform binomial dependent variables such as mine (Jaeger, 2008). Second, they allow the inclusion of continuous predictor variables, just like traditional regression analyses. Thus, continuous variables do not need to be discretised, which would lead to a loss of statistical power (Cohen, 1983). Third, they are able to deal with unbalanced data sets (see Baayen et al., 2008), though this advantage is less relevant in the present study. Fourth, they allow for the joint modelling of participant-related and item-related effects. These properties make mixed models eminently well-suited as the tool of analysis for my present purposes.

I will not describe in detail the formal underpinnings of linear mixed-effects models. Instead I refer to Baayen (2008), Baayen et al. (2008) and Jaeger (2008) for introductions geared towards language researchers and to Faraway (2006) and Zuur et al. (2009) for more technical accounts. Informally, mixed effects models describe the outcome variable as a function of *fixed effects*, which can loosely be defined as effects that are expected to hold across participants and items, on the one hand and by-participant and by-item adjustments to the predicted outcomes (called *random effects*) on the other hand. By including random effects for participants and items, researchers automatically specify their data's dependency structure, thereby circumventing the independence assumption of traditional models. The by-participant adjustments, which are drawn from a normal distribution with a mean of 0 and unknown variance  $\sigma^2$ , do justice to the commonsensical realisation that some participants score better or worse on a task due to factors not yet modelled or due to reasons unknown: participants with negative by-participant adjustments score worse than would be expected on the basis of the fixed effect predictions; participants with positive by-participant adjustments score better. Similarly, the by-item adjustments, also drawn from a normal distribution with a mean of 0 and unknown variance, statistically account

for the fact that some items are easier or harder to process due to factors not under the researchers' control.

The by-participant and by-item adjustments that I have just described are more aptly referred to as *random intercepts* as they shift the overall individual modelled predictions up- or downwards. When dealing with dependent data points, however, there is a further factor to consider: some participants may be more responsive to certain stimulus characteristics and some items may be more strongly or weakly affected by particular participant characteristics. To stay within the realm of receptive multilingualism research, the number of languages known may not affect the intelligibility of internationalisms as strongly as that of more language-specific stimuli. Similarly, longer words may be more intelligible than shorter words generally, but the effect might not be as large in children.<sup>14</sup> In extreme cases, overall trends in the data may even be reversed in some dependency clusters, which is referred to as *Simpson's (1951) paradox* (for a linguist's introduction to Simpson's paradox, see Jaeger et al., 2011; see also Kievit et al., 2013). Fortunately, such by-participants and by-item adjustments can—and indeed should, see Barr et al. (2013), Jaeger et al. (2011) and Schielzeth and Forstmeier (2009)—be modelled as well. They are referred to as *random slopes*. Fixed effects that remain robust after the inclusion of their associated random effects can reasonably be expected to generalise to new participants and items, i.e. they are prime candidates for population-wide effects. That is not to say that the by-item and by-participant adjustments are in themselves uninteresting, however, as their specific patterning may require an explanation and prompt new research.

To round off this introduction to mixed models, I point out that linear mixed model outputs do not come with a default  $R^2$  coefficient of determination that readers versed in the traditional linear regression model will be familiar with and neither do logistic mixed model algorithms dutifully report pseudo- $R^2$  indices such as Nagelkerke's. The issue of defining such measures for (generalised) mixed models is non-trivial due to the more complex specification of these models, and while some algorithms have been proposed to compute mixed model  $R^2$  pendants, these algorithms are neither straightforwardly implementable nor as of yet in widespread use. Therefore, instead of absolute (pseudo-) $R^2$  measures, relative goodness-of-fit measures such as the AIC (Akaike,

---

<sup>14</sup>Both hypothesised effects have been concocted solely for illustrative purposes.

1974) will be used in order to compare the fit of different models in this thesis. The AIC measure provides a numerical estimate of the model fit–model complexity trade-off. More complex models, i.e. models with more predictors or interactions, will necessarily fit the data better in absolute terms but run the risk of modelling relatively more ‘noise’, i.e. random fluctuations in the data. The AIC measure tells us whether increased model complexity is accompanied by a sufficiently large increase in the fit to the data to warrant the inclusion of the additional model parameters: the lower the AIC measure is, the better the model is relative to the models with which it is compared. Alternatively, log-likelihood ratio tests can be used to compare mixed model fits.

The generalised linear mixed-effects models in this thesis were fitted with the `lmer()` function in the `lme4` package (version 0.999999-2; Bates et al., 2013) for R (R Core Team, 2013).

### 4.3.2 Generalised additive models

Like generalised linear models (GLMs), generalised linear mixed-effects models assume that the relationship between the predictor covariates and the outcome variable is approximately linear. In the case of logistic models, the relationship needs to be roughly linear in log-odds space. GLMs and GLMMs can accommodate for non-linearities to some extent by means of modelling the covariate–outcome relationship as a higher-order polynomial or by cubic spline fitting, but a more principled way of dealing with non-linearities is to make use of *generalised additive models* (GAMs).

Like GLMs, GAMs can model non-Gaussian outcome variables, such as the binary accuracy variable in this case, in terms of several predictor variables. However, the requirement of generalised linear models that the relationships between the outcome and the predictors be linear is relinquished in generalised additive models. Instead, non-linear relationships can be modelled, the form of which is estimated from the data. This is essentially accomplished by fitting higher-order polynomial regressions on subsets of the data and glueing the pieces together. The more subset regressions are fitted and glued together, the more ‘wiggly’ the overall curve will be. Fitting too many subset regressions results in overwiggly curves that fit disproportionately much noise in the data (‘oversmoothing’). In order to prevent this, the algorithm can be furnished with a cross-validation procedure, or an generalised (algebraic) approximation

thereof (Wood, 2006): oversmoothed curves will perform poorly in cross-validation, and the cross-validation procedure will identify the number of subset regressions, and hence the form of the overall curve, that has the best chance of predicting new data points. It is well beyond the scope of this dissertation to discuss the fine points of generalised additive modelling, and I refer the interested reader to Chapter 3 in Zuur et al. (2009) for a technical but accessible introduction.

GAMs can be supplemented with random effects in order to cope with dependency structures (see Section 4.3.1), giving rise to generalised additive mixed-effect models (GAMMs). The non-linear counterparts to random slopes, *factor smooth interactions*, representing for instance by-participant or by-item discrepancies from an overall wiggly curve, could not be fitted in models with crossed dependency structures (e.g. items and participants) at the time of writing. Random intercepts and random slopes for linear effects could be fitted, however. All in all, then, GAMMs are attractive tools when the data is characterised by a strong degree of non-linearity, but when the relationships are only mildly non-linear, I deem it preferable to model the data linearly and make use of the facility to model random slopes.

To conclude, a word on interpreting the output of a GAM or GAMM. The algorithms produce a wealth of numerical information, including significance tests that are computed with respect to a number referred to as the ‘estimated degrees of freedom’. This number represents the wiggleness of the effect that was modelled non-linearly: estimated degrees of freedom near 1 indicate that the effect is essentially linear, higher values indicate a higher degree of wiggleness. However, the functional form of the relationships between the predictors and the outcome can only be inferred by inspecting the models’ visual output. Moreover, the models’ algorithms dutifully produce  $p$ -values based on  $F$ -tests, but these are intended as approximations rather than as 100 % accurate numbers. In Wood’s (2006) own words, “it is usually better to be able to say something approximate about the right model, rather than something very precise about the wrong model” (p. xvii). The implication is two-fold. First, the traditional strict boundary between significant and non-significant  $p$ -values at  $\alpha = 0.05$  is even less likely than usual (see e.g. Cohen, 1994; Gigerenzer, 2004) to be a useful cut-off in GAMs and GAMMs. Second, new software versions may yield different, presumably more accurate estimates. In sum, strict reliance on numerical summaries

and sharp cut-off marks are even less desirable in generalised additive modelling than when using more established methods.

The generalised additive mixed-effects models in this thesis were fitted with the `bam()` function in the `mgcv` package (version 1.7-24; Wood, 2013) for R (R Core Team, 2013).



# Chapter 5

## Data inspection

This chapter presents preliminary analyses on the dependent and independent variables with a view towards identifying and rectifying potentially problematic data patterns. The multivariate modelling of the data is then carried out in Chapters 6 and 7.

### 5.1 Cognate guessing data

Four of the 167 participants were not able to complete the cognate guessing task owing to computer malfunctions. The results of the 163 remaining participants are summarised in Table 5.1 on the next page. Importantly, the cognate guessing task does not appear to have been overly easy or difficult, as evidenced by the lack of floor and ceiling effects for the outcome variables of interest, viz. the number of correctly translated target stimuli in the two modalities.

Table 5.1 also presents summary data for the number of correctly translated profile words. These are everyday words that should be indecipherable to readers or listeners without any knowledge of Swedish or another North-Germanic language. I included them in order to identify participants with such knowledge and exclude them from the data set. None of the participants were able to correctly translate more than two profile words in either modality, however, and therefore no participants were excluded from the analyses on the grounds of having substantial prior knowledge of the *Lx*. The fact that some participants managed

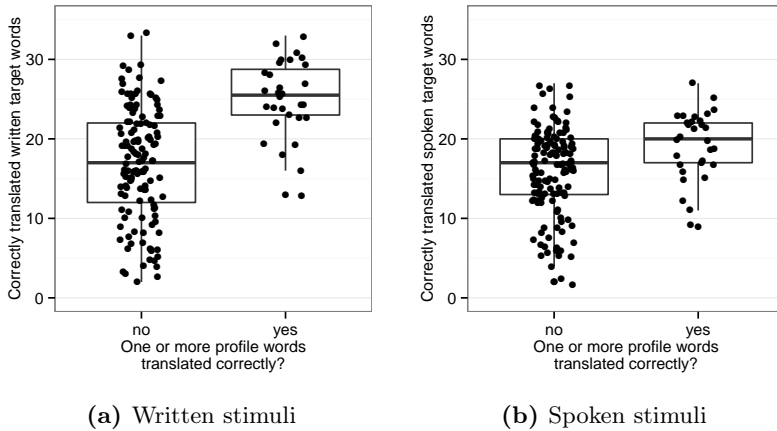
**Table 5.1:** Summary data for the number of correctly translated stimuli per participant in the cognate guessing task.

		Max	Range		Median	Mean	SD
			Lower	Upper			
Written	Target stimuli	45	2	33	19	18.5	7.4
	Profile stimuli	5	0	2	0	0.1	0.3
Spoken	Target stimuli	45	2	27	17	16.5	5.4
	Profile stimuli	5	0	2	0	0.2	0.4

to translate some profile words correctly can most likely be ascribed to a small degree of incidental learning, e.g. during holidays or due to popular culture. The profile word that was translated correctly most often (21 times) is [ˈsfærjɛ] (*Sverige* ‘Sweden’), followed by [ˈɛlska] (*älska* ‘to love’) and *barn* ‘child’ (both 5 times).

Though no participants were excluded for having prior knowledge of Swedish, it is possible that prior incidental learning is associated with a higher translation success rate. In Figure 5.1, I plotted the number of correctly translated target words for the participants that were able to translate at least one profile word ( $n = 30$ ) versus for those that were not ( $n = 133$ ). Panel (a) shows that participants who translated at least one profile word correctly scored indeed better on the written target words ( $M = 25.1$ ,  $SD = 4.8$ ) than those who did not ( $M = 17.0$ ,  $SD = 7.0$ ). Similarly, panel (b) shows those who were able to translate at least one profile word correctly translated more spoken target words correctly on average ( $M = 19.4$ ,  $SD = 4.2$ ) than those who were not ( $M = 15.9$ ,  $SD = 5.4$ ). It thus seems that prior incidental learning, as rudimentarily assessed by means of profile word translations, is associated with higher translation accuracy on target words. In order to account for this possibly confounding effect in my analyses, I created a binary variable that indicates whether a participant had been able to translate at least one profile word correctly. This indicator variable aside, my analyses will be concerned only with the target stimuli.

As for the target stimuli, three aurally presented words were not translated correctly by any of the participants: [ˈfɛ:ɣɛl] (*kägel* ‘cone’), [ˈty:dliɡ] (*tydlig* ‘clear’) and [ʃærm] (*skärm* ‘screen’). Since these items had no discriminatory power and since items with no response variation



**Figure 5.1:** Boxplots of the number of correctly translated target words per participant in the written and in the spoken modality according to whether the participants were able to translate at least one profile word correctly ( $n = 30$ ) or not ( $n = 133$ ).

tend to cause convergence issues in the modelling stage, they were discarded from the analyses, leaving 87 items for 163 participants for a total of 14,181 data points. None of the target words were translated correctly by all of the participants. The full-fledged item analysis is postponed to Part III.

In addition to measuring translation accuracy, I extracted the time that it took the participants to decide whether to attempt a translation (measured from stimulus onset onwards). These response latencies served a filtering purpose: unrealistically fast responses are likely due to inadvertent key strokes and can be identified and removed from the dataset. I discarded one data point for a written stimulus that was associated with a latency of less than 250 ms and a further seven for spoken stimuli for which the latencies were faster than the duration of the stimuli. All of these eight data points were blank responses. Removing them left a total of 14,173 data points.

## 5.2 Inspection of the linguistic and cognitive variables

### 5.2.1 Self-assessed language skills and number of foreign languages

When asked to provide CEFR-based self-ratings, a substantial minority of our German-speaking participants indicated that their receptive standard German skills were not sufficiently good to conform to the description of the C2 level ('mastery'): 48 out of 161 participants provided a self-assessment of C1 or lower for their reading skills (2 blank responses) and 39 out of 160 participants did so for their listening skills (3 blank responses). This cannot entirely be due to their having misunderstood the instructions and flipping the rating scale as only 5 and 6 participants provide self-ratings of A1 for reading and listening, respectively. Neither can it be entirely due to their overall modesty as some give higher self-assessments for their French or English skills than for German. Possibly, these lower-than-expected self-ratings are linked to a common belief among Swiss Germans, namely that their own variety of standard German (*Schweizerhochdeutsch*) is less correct than the variety spoken in Germany (Scharloth, 2006), but it is not clear why this would affect the self-ratings for their receptive skills. All in all, the self-ratings for German cast doubt on the reliability of the self-ratings for the other languages in the participants' repertoires, too. Consequently, I will not use the self-assessed language skills in my analyses.

A variable I will take into consideration is the number of languages other than Swiss German dialects and standard German that the participants listed in the questionnaire. This tally includes neither dead languages, notably Latin and Ancient Greek, nor sign languages. Only six participants listed more than five languages. In order to prevent these participants from exerting undue influence on the analyses, I collapsed them into the same category as participants with five foreign languages. The languages listed were the following: English (159), French (148), Italian (85), Spanish (57), Portuguese (7), Tagalog (5), Serbian (4), Hungarian, Romansh (both 3), Arabic, Cebuano, Dutch, Greek, Russian, Swahili (all 2), Bahasa, Catalan, Czech, Hebrew, Romanian, Tamil, Telegu, Thai and Turkish (all 1). Thus, only two participants had some minimal knowledge of a Germanic language other than English,

**Table 5.2:** Summary data for the participant-related linguistic and cognitive measures.

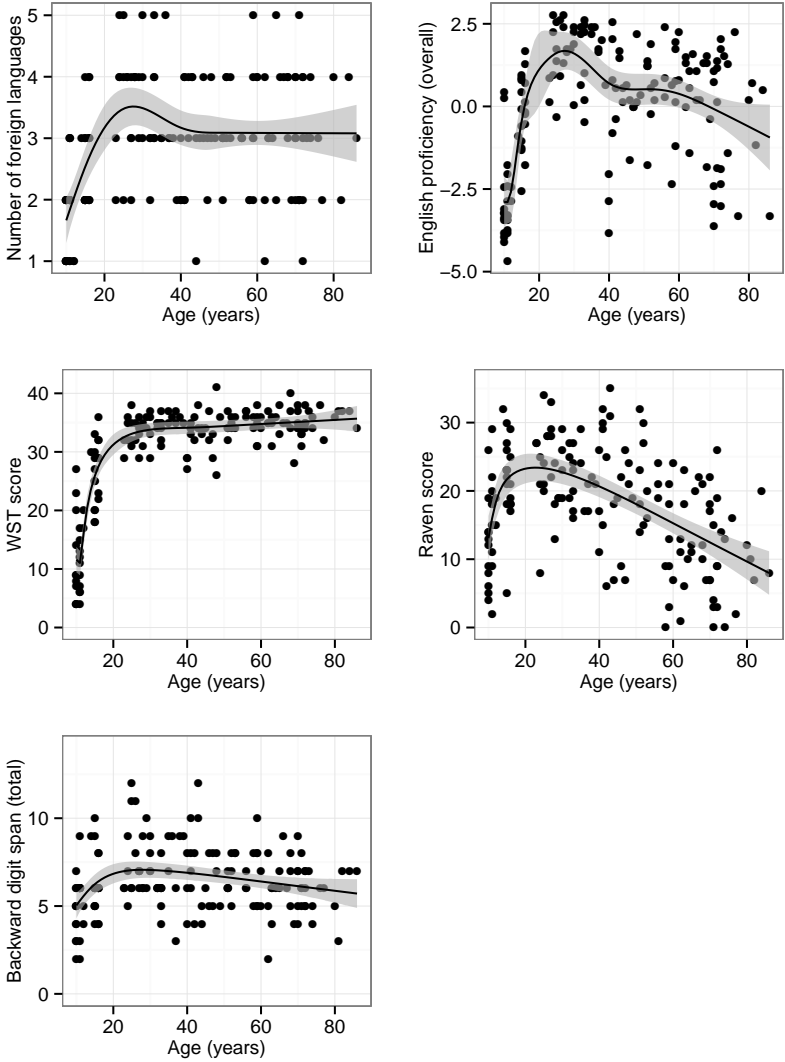
		<i>n</i>	Range		Median	Mean	SD
			Lower	Upper			
Number of foreign languages		163	1	5	3	3.0	1.1
English test	Multiple choice	160	0	20	15	13.7	4.6
	C-test	160	0	24	15.5	14.6	6.8
	Overall <sup>a</sup>	160	-4.7	2.8	0.5	0.0	1.9
WST		162	4	41	34	30.2	8.8
Raven		163	0	35	19	17.8	8.2
BW-DS	Span	163	2	8	4	4.6	1.2
	Total	163	2	12	6	6.4	1.9

<sup>a</sup> The overall English score is the sum of the *z*-normalised scores on the multiple choice test and the C-test.

viz. Dutch. Table 5.2 presents summary statistics for this variable and the upper left panel of Figure 5.2 on the following page sketches its sample lifespan trajectory. Note that, in this sample, the number of foreign languages known remains roughly stable throughout the adult lifespan, but that it increases throughout childhood and adolescence as a result of formal schooling.

## 5.2.2 English language proficiency test

Two tasks were used to measure the participants' proficiency in English: a multiple choice task and a C-test. The scores on these two tasks are unsurprisingly strongly correlated ( $r = 0.85$ ), for which reason I combined them into a single 'overall' measure. In order to weight the two subtask scores equally in this composite measure, I *z*-normalised them, i.e. centred them at their means and divided them by their respective standard deviation, before adding them up for each participant. Doing so guarantees that the composite measure correlates equally strongly with both underlying measures. The new overall English proficiency measure correlates at  $r = 0.96$  with both original measures. Table 5.2 presents the relevant summary statistics of both the overall and the original measures, and the upper right panel of Figure 5.2 shows how the overall English proficiency measure varies as a function of age in the participant sample. Note that this measure reaches its zenith around age 30 and decreases from that point onwards.



**Figure 5.2:** Sample age trends in the participant-related predictors.

### 5.2.3 German vocabulary test

Summary data for the raw scores on the German vocabulary task (*Wortschatztest*, WST) are given in Table 5.2. Their lifespan development is plotted in the middle left panel of Figure 5.2 and is roughly what one would expect of measure of crystallised intelligence: it increases sharply up until about age 30 and remains stable from that point onwards till old age.

Despite explicit instructions to the contrary, many participants clearly guessed when they did not know the right answer. This is demonstrated by the fact that 29 participants gave more than five wrong non-blank responses. Adjusting for guessing by subtracting the number of incorrect non-blank responses from the number of correct responses resulted in a measure that is very strongly correlated with the uncorrected measure ( $r = 0.93$ ), but which is extremely strongly left-skewed. In what follows, I will consequently make use of the uncorrected WST measure.

### 5.2.4 Raven's advanced progressive matrices

Table 5.2 presents summary data for the raw scores on the Raven task. The sample age trend for this measure, plotted in the middle right panel of Figure 5.2, conforms to the canonical lifespan trajectory of fluid intelligence.

### 5.2.5 Backward digit span task

The backward digit span task (BW-DS), which served as a working memory (WM) task, produced two measures: a 'span' measure representing the length of the longest digit sequence that the participant could repeat backwards and a 'total' measure representing the total number of correct responses. Summary data for both are presented in Table 5.2. These two measures are necessarily highly correlated ( $r = 0.92$  in this sample), which is why I settled on the more fine-grained 'total' measure as the WM indicator for my analyses.

The sample age trend for the BW-DS 'total' measure is shown in the bottom panel of Figure 5.2. While the measure shows some increase up until age 30, it does scarcely reflect the canonical age-related decline in working memory in older participants. Speculatively, the fact that the sample was self-selected may go some way in accounting for the

absence of a strong age effect in WM: older participants with smaller WM capacity may simply have been less likely to sign up for a 3-hour data collection session. Alternatively, or additionally, the BW-DS may have been too crude a measure of WM to grasp a stronger, present age-related decline: the BW-DS is a rather quick-and-dirty task for the purposes of tapping working memory that is considered by some (e.g. Engle et al., 1999; Park and Payer, 2006) to be a test of short-term memory, which is less affected by ageing than is working memory (Park and Payer, 2006), instead (see Section 4.2.4).

### 5.3 Missing data imputation

As Table 5.2 reveals, English proficiency data are available for only 160 out of 163 participants. On top of that, there are no WST data for one additional participant, leaving 159 participants with the complete set of relevant linguistic and cognitive measures. These data are missing due to experimenter error and can be assumed to be missing at random, i.e. their missingness is not a function of the true values of the missing data.

Since the algorithms used for statistical analysis perform listwise deletion when the set of predictor variables is incomplete, I would effectively be giving up four sets of perfectly valid data points. I therefore decided to salvage these data points by imputing the three missing English proficiency values as well as the missing WST score. To this end, I used the *Amelia* package (version 1.6.4; Honaker et al., 2012) for R. Rather than substituting the missing values with a typical univariate value, e.g. their means, the algorithms in *Amelia* take into account multivariate relationships, i.e. they impute the missing values of one variable given what is known about the correlations between this variable and other variables.

The following variables were entered into the imputation model: (a) the number of correctly translated written target words; (b) the number of correctly translated spoken target words; (c) the binary variable indicating whether the participant had been able to translate at least one profile word correctly; (d) the participant's sex; (e) their age; (f) the number of foreign languages in their repertoire; (g) WST score (one value to be imputed); (h) Raven score; (i) BW-DS total score; and (j) the overall score for English proficiency (three values to be imputed).

(a) and (b), i.e. the dependent variables of the analyses, were included in the imputation model as recommended by Graham (2009) and Honaker et al. (2012, package vignette). The data were imputed only once.<sup>15</sup> The analyses reported in the remainder of this thesis are based on this imputed data set.<sup>16</sup>

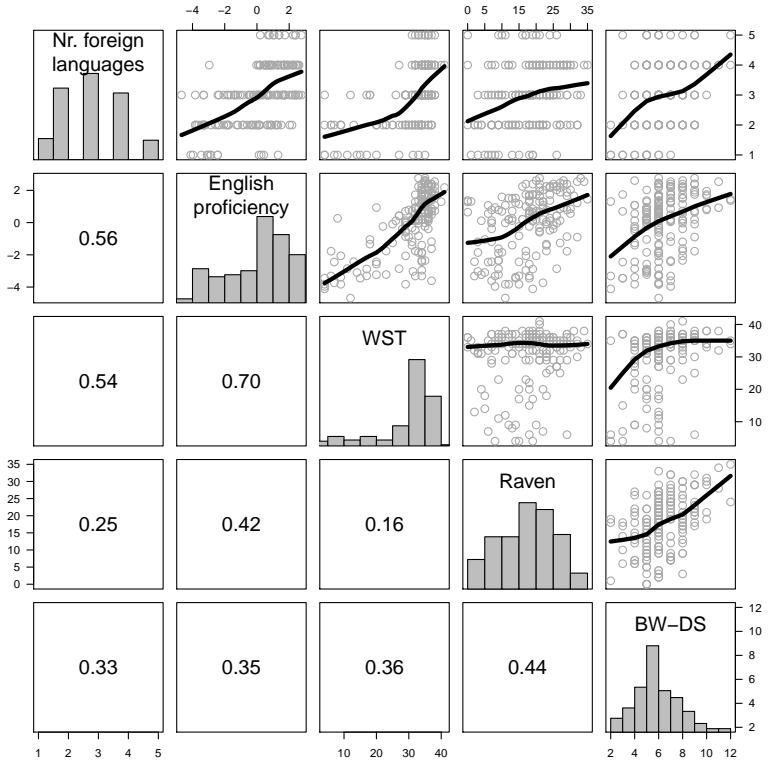
## 5.4 Multicollinearity assessment

The linguistic and cognitive participant-related variables are unsurprisingly intercorrelated to a certain extent: as Figure 5.3 on the next page shows, participants who know many foreign languages tend to perform well on the English proficiency and cognitive tests and *mutatis mutandis* for the other variables. Substantial multicollinearity can make it difficult to gauge the influence of any one predictor in a regression model on the outcome variable. However, a numerical check revealed that the degree of multicollinearity in this set of predictor covariates can be considered to be negligible when they are properly centred at their means ( $\kappa = 3.4$ ;  $\kappa$  values of 6 or less indicate a negligible degree of collinearity, see Baayen, 2008, p. 182). I therefore refrained from eliminating the intercorrelations in the data set entirely, e.g. by means of principal component analysis or residualisation, as I felt that in the present case, this would render the predictor variables difficult to interpret.

---

<sup>15</sup>The *Amelia* package also supports multiple imputation, i.e. the creation of several parallel data sets in which the imputed values differ somewhat, thereby reflecting the uncertainty regarding their precise values. The differences between the parallel data sets are introduced by the bootstrapping algorithm involved, which relies on random sampling. These parallel data sets can then be subjected to the same set of analyses, with differences in the outcomes of these analyses again reflecting the uncertainty about the missing values. GAMM modelling in itself is already computationally rather expensive, however; running the same analysis on several parallel data sets compounds this drawback. Furthermore, whereas regression coefficients produced by several linear regression models can straightforwardly be averaged, it is not trivial to average over the output of several additive models. Lastly, the degree of missingness in my data is quite limited and it is unlikely that my results or their statistical significance critically hinge on the imputed data. I therefore decided to impute the missing data only once rather than to perform multiple imputation.

<sup>16</sup>For papers on this project, I only analysed the data for the 159 participants for whom all relevant data were available. The models in these papers are highly similar to the ones reported here and allow the same conclusions to be drawn.



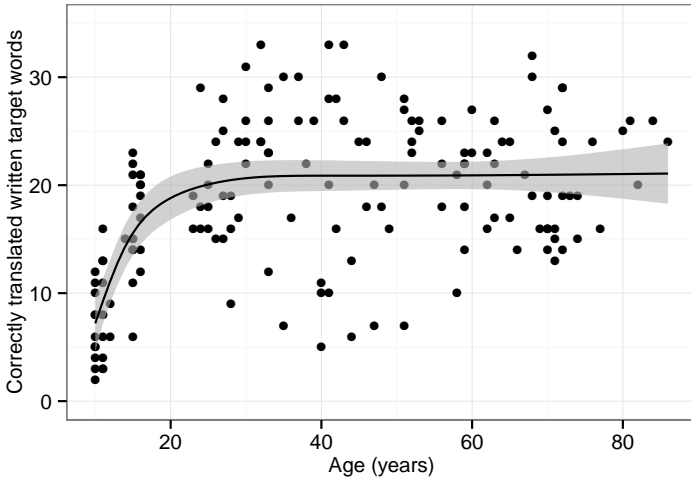
**Figure 5.3:** Intercorrelations between the participant-related linguistic and cognitive predictors. Upper triangle: Bivariate scatterplots with non-parametric scatterplot smoothers. Main diagonal: Histograms. Lower triangle: Pearson correlation coefficients.

## Chapter 6

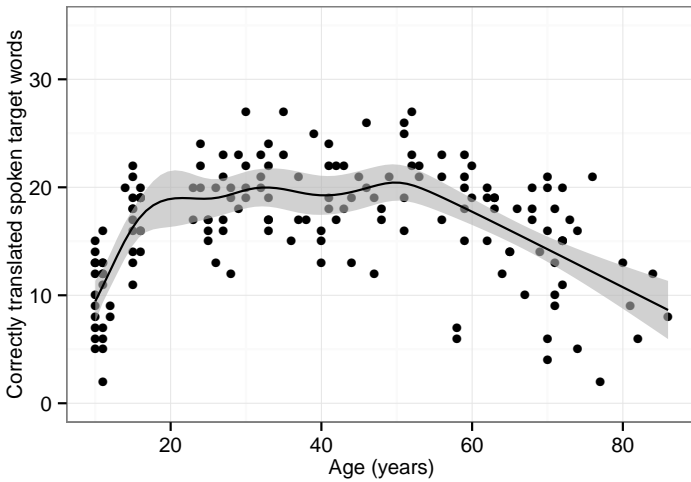
# Age trends in cognate guessing success

The primary research question in the first part of this thesis concerns the lifespan development of cognate guessing skills in the written and spoken modalities. Even though the tallies of correctly translated spoken and written target words per participant are substantially correlated ( $r = 0.56$ ), Figure 6.1 on the following page suggests that there are differences in the age-patterning between the two modalities: performance in both modalities increases up until about age 20–25, but whereas cognate guessing performance in the written modality remains stable in the following decades, it remains stable only up until roughly age 50 and decreases from that point onwards in the spoken modality.

This visual exploration suffers from two drawbacks, however. First, it is based on the number of correctly translated answers per modality per participant. This means that the individual cognates are not treated as samples from a larger population, which would be preferable for statistical and conceptual reasons as outlined in Section 4.3. Second, this bivariate analysis ignores the effects of potential confound variables. The first obvious potential confound variable is the participants' sex. Table 4.1 showed that the two sexes are not similarly distributed in the different age groups. To the extent that men perform differently from women, this unequal distribution may affect the overall age trends in Figure 6.1. The second potential confound variable is whether the



(a) Written stimuli



(b) Spoken stimuli

**Figure 6.1:** Number of correctly translated target words per participant as a function of age ( $n = 163$ ).

participants translated at least one profile word correctly: participants who did tended to translate more target words correctly as well (see Figure 5.1), but none of the participants aged 20 or younger belonged to this group.

For these reasons, I verified the age trends in a multivariate mixed-effects model. Since the relationship between age and translation accuracy is clearly non-linear, I fitted binomial generalised additive mixed effects models (GAMMs) with crossed random intercepts for participants and items. Ideally, I would have liked to also take into account by-item differences in the age effects. As of yet, however, this is not possible in GAMMs with crossed random effects (see Section 4.3.2). In a first step, I fitted separate models for the written ( $n = 7,334$ ) and for the spoken items ( $n = 6,839$ ). These models included a non-linear term for age as well as parametric terms for the participants' sex and for the binary variable indicating whether the participant provided at least one correct profile word translation.<sup>17</sup> The participants' sex did not contribute significantly to the fit of either model and was consequently removed from both models; the effect of profile word translation was not significant in the spoken-item model and was likewise discarded from this model. The two models are summarised numerically in Tables 6.1 and 6.2 on page 81; the models' age terms are plotted in Figure 6.2.

When comparing Figures 6.1(a) and 6.2(a), it can be seen that the multivariate approach reveals that written target word translation success may not be wholly stable throughout the adult lifespan. Rather, Figure 6.2(a) suggests that cognate guessing performance on written items may even increase somewhat with age, even in adults, when the confound effect of correct profile word translation is taken into account. As for the spoken items, Figure 6.2(b) presents a picture similar to Figure 6.1(b).

Lastly, I verified whether the diverging age trends for written and spoken cognate guessing are indeed statistically reliably different from one another. To this end, I fitted two additional GAMMs. In both models, the accuracy of all 14,173 responses was modelled in terms of stimulus modality and a non-linear term for age. The binary variable

---

<sup>17</sup>As a technical aside, the non-linear age term was fitted using an adaptive smooth rather than thin plate regression splines (the default) or cubic regression splines (a common alternative). Adaptive smooths do not assume that the data are uniformly wiggly along the covariate, and as Figure 6.1 shows, there is hardly any wiggleness for certain stretches on the age covariate.

**Table 6.1:** GAMM modelling translation accuracy on written target words in function of age. Panel (a): Parametric fixed effects, their standard errors and their significance. Panel (b): Smooth term with its estimated degrees of freedom,  $\chi^2$ -statistic and significance. Panel (c): Modelled standard deviation of the random effects ( $\hat{\sigma}$ ). Parameter estimates are expressed in log-odds.

(a) Parametric terms

	Estimate $\pm$ SE	$p$
Intercept	$-0.80 \pm 0.28$	0.005
$\geq 1$ correct profile word translation	$0.90 \pm 0.18$	$<0.001$

(b) Smooth term

	Est. df	$\chi^2$	$p$
Age	3.4	99.9	$<0.001$

(c) Random effects

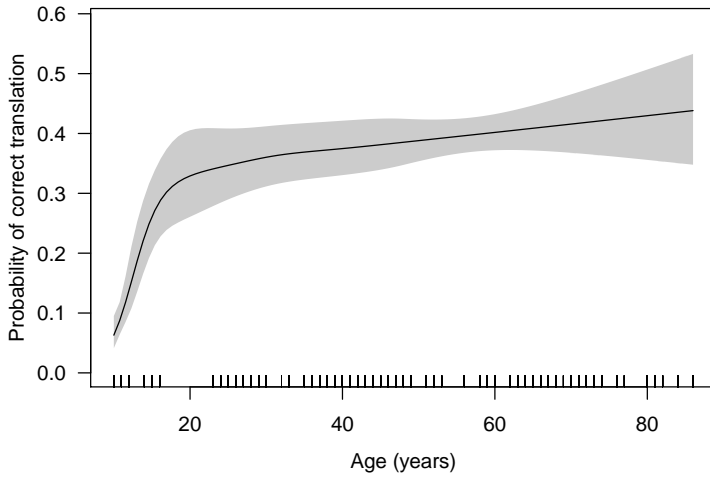
	$\hat{\sigma}$
Random intercept by participant	0.75
Random intercept by items	1.8

**Table 6.2:** GAMM modelling translation accuracy on spoken target words in function of age. (a): Parametric fixed effect, its standard error and its significance. (b): Smooth term with its estimated degrees of freedom,  $\chi^2$ -statistic and significance. (c): Modelled standard deviation of the random effects ( $\hat{\sigma}$ ). Parameter estimates are expressed in log-odds.

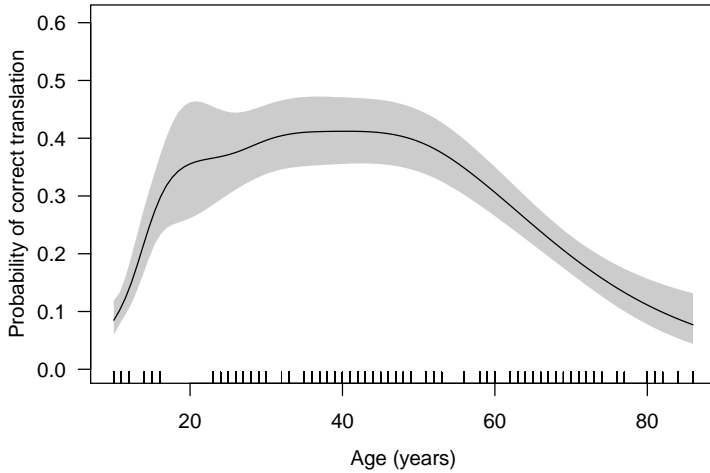
(a) Parametric term			
	Estimate $\pm$ SE		$p$
Intercept	-1.0 $\pm$ 0.4		0.006

(b) Smooth term			
	Est. df	$\chi^2$	$p$
Age	4.9	114.8	<0.001

(c) Random effects	
	$\hat{\sigma}$
Random intercept by participant	0.66
Random intercept by items	2.4



(a) Written stimuli



(b) Spoken stimuli

**Figure 6.2:** Age trends in GAMM-fitted translation accuracy with 95% confidence bands.

indicating whether the participant provided at least one correct profile word translation was included as well, as was the participants' sex. In the first model, stimulus modality was allowed to interact with sex and profile word translation but not with the non-linear age term. In the second model, stimulus modality was allowed to interact with age as well. If the age trends differ significantly between the modalities, the model with the age  $\times$  modality interaction should fit the data better than the model without this interaction. The AIC value of the second model was indeed substantially lower (i.e. better) than that of the first ( $\Delta$  AIC = 142), indicating that the age trends are indeed reliably different from one another in the two modalities. The output of these models is not provided here as it does not yield any insights not yet provided by the models summarised in Tables 6.1 and 6.2.

To summarise, there are diverging age trends for *Lx* cognate guessing success in the written and in the spoken modality: *Lx* cognate guessing success appears to increase steeply throughout childhood and adolescence in both modalities, but whereas it seems to remain stable or even show further increase throughout the adult lifespan in the written modality, it declines from roughly age 50 onwards in the spoken modality. In the next chapter, I explore how these diverging age trends can be explained in terms of linguistic and cognitive factors.



## Chapter 7

# The impact of language skills and cognitive characteristics

I now turn to the statistical modelling of cognate guessing accuracy in terms of the participants' linguistic and cognitive data. In order to keep this presentation tractable, I fitted separate models for the written items and for the spoken ones rather than one all-inclusive model with several by-modality interaction terms.

### 7.1 Written items

7,334 valid responses to written stimuli were analysed. Following Baayen (2008, pp. 254–255), all continuous predictors were centred at their sample means. Exploratory GAMM-based analyses (not reported here) did not reveal any substantial non-linear patterns between the predictor covariates and the outcome variable in log-odds space. The data were therefore further modelled in a generalised linear mixed model. The effects of sex, Raven score and backward digit span were small and non-significant and were dropped from the model. Doing so did not appreciably affect the parameters and the standard errors of the effects remaining in the model.

The model's fixed effects are summarised numerically in Table 7.1(a); a visual summary is provided in Figure 7.1 on page 88. In deference to readers unfamiliar with logistic regression outputs, I will walk through the model step by step to demonstrate how the probabilities in Figure 7.1 can be derived from the parameters reported in Table 7.1(a).

The intercept value in Table 7.1(a),  $-0.93$ , represents the model's estimate of the probability with which a randomly chosen participant who did not provide any correct profile word translations and whose other covariate measures correspond to the sample means can correctly translate a randomly chosen target word. This estimate is expressed in log-odds, or logits. The relationship between an actual probability  $p$  and its logit,  $\text{logit}(p)$ , is defined as follows:

$$\text{logit}(p) \equiv \ln \frac{p}{1-p}$$

where 'ln' stands for the natural logarithm, i.e. the logarithm with base  $e \approx 2.718$ . To derive  $p$  from  $\text{logit}(p)$ , we need to apply the inverse of the logit function, the logistic function:

$$p = \frac{e^\alpha}{1 + e^\alpha}$$

where  $\alpha \equiv \text{logit}(p)$ . For  $\alpha = -0.93$ , as in the present case, this gives  $p \approx 0.28$ . If we want to arrive at the modelled probability with which a randomly chosen target word can be translated correctly by a randomly chosen participant whose covariate measures correspond to the sample mean and who *did* translate at least one profile word correctly, we need to add the estimate for ' $\geq 1$  correct profile word translations' from Table 7.1(a) to  $\alpha$ :  $-0.93 + 0.50 = -0.43$ , which gives  $p = \frac{e^{-0.43}}{1 + e^{-0.43}} \approx 0.39$ . Thus, participants who translated at least one profile word correctly are more likely to translate a random target word correctly. The upper left panel of Figure 7.1 presents this 'partial' fixed effect in probabilities. (Note, however, that for the partial effects in Figure 7.1, the continuous variables were fixed at their medians, not at their means, hence the difference between the plotted estimates and the values that we have just calculated.)

Turning to the predictors of interest, Table 7.1(a) reveals a fixed effect of the number of foreign languages in the participants' repertoires that is significant at the 0.05 threshold. To express the effect in probabilities,

**Table 7.1:** Generalised (logistic) mixed-effect model modelling translation accuracy on written target words in function of participant-related linguistic and cognitive predictors. Panel (a): Fixed effects, their two-tailed significance and their effect sizes. Panel (b): Modelled standard deviation of the random effects ( $\hat{\sigma}$ ). All continuous variables were centred at their sample means. Parameters and effect sizes are expressed in log-odds and are reported to two significant digits.

(a) Fixed effects

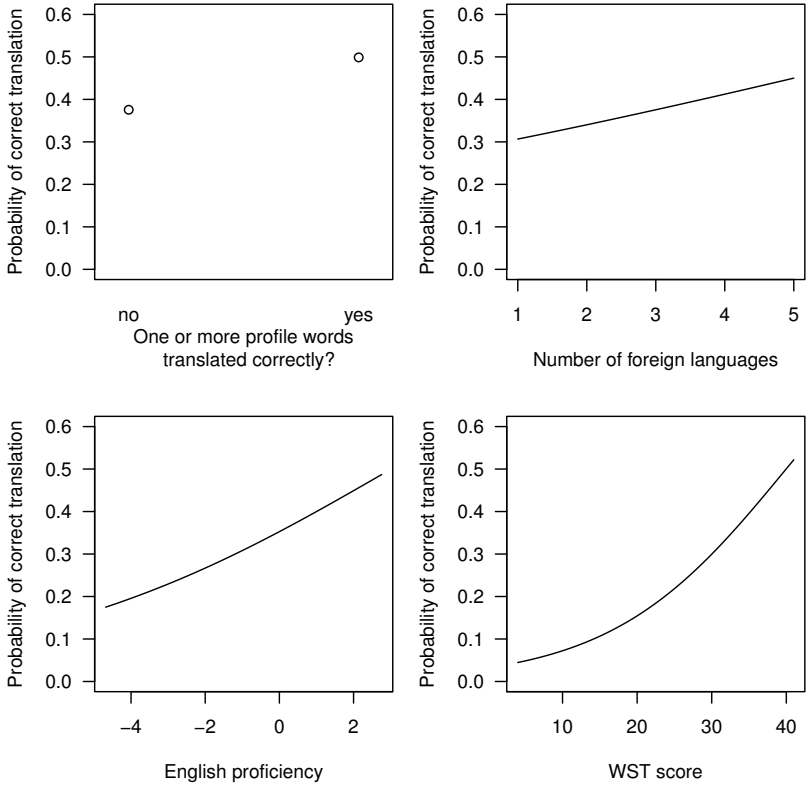
	Estimate $\pm$ SE	$p$	Effect size $\pm$ SE <sup>a</sup>
Intercept <sup>b</sup>	-0.93 $\pm$ 0.32	0.003	
$\geq 1$ correct profile word translations	0.50 $\pm$ 0.17	0.003	0.50 $\pm$ 0.17
Number of foreign languages	0.15 $\pm$ 0.07	0.036	0.62 $\pm$ 0.29
English proficiency	0.20 $\pm$ 0.06	<0.001	1.5 $\pm$ 0.4
WST score	0.085 $\pm$ 0.015	<0.001	3.2 $\pm$ 0.6

<sup>a</sup> Following Baayen et al. (2008), effect sizes were computed as the largest absolute difference in the outcome variable (in log-odds) when the predictor variable is allowed to vary along its range. For instance, the centred English proficiency measures spans from  $-4.7$  to  $2.8$ . Since the parameter estimate for this variable is  $0.20$ , its effect size equals  $0.20 \times |2.8 - (-4.7)| \approx 1.5$ .

<sup>b</sup> The intercept represents the predicted probability of a correct target word translation by a randomly chosen participant who did not provide any correct translation to a profile word and whose other covariate values correspond to the sample means.

(b) Random effects

	$\hat{\sigma}$
Random intercept by participant	0.68
Random intercept by item	2.0
Random slope for English proficiency by item	0.21
Random slope for WST score by item	0.061



**Figure 7.1:** Partial fixed effects of the GLMM modelling translation accuracy on written target items in terms of participant-related linguistic and cognitive predictors. Nominal variables not included in a given plot were fixed at their modes, and continuous variables not included in a given plot were fixed at their medians.

let us compute the probability with which randomly chosen target word can correctly be translated by a randomly chosen participant who did not provide any correct profile word translations, who knows five foreign languages (two more than the sample average) and whose other covariate measures correspond to the sample mean. This participant's  $\alpha$  is  $-0.93$  (the intercept)  $+ 2 \times 0.15$  (the estimated parameter for 'number of foreign languages'), which equals  $-0.63$ . The probability we want to calculate is therefore  $p = \frac{e^{-0.63}}{1 + e^{-0.63}} \approx 0.35$ , i.e. about 7 percentage points higher than for someone with an average-sized linguistic repertoire. The upper right panel of Figure 7.1 presents this partial effect.

The effects of the other covariates can similarly be gleaned from Table 7.1(a) and Figure 7.1. (Note that the plotted partial effects for these covariates are not perfectly straight lines. This is the result of the logistic transformation: the effects are modelled linearly in terms of log-odds, but transforming them back into actual probabilities results in curved trend lines.) Thus, in terms of effect sizes (ES), the participants' English proficiency seems to be a more important linguistic factor contributing to cognate guessing accuracy (ES:  $1.5 \pm 0.4$ ) than the number of foreign languages in the participants' repertoires (ES:  $0.62 \pm 0.29$ ). The most important predictor of all, however, is the score on the German vocabulary test (WST), a measure of crystallised intelligence (ES:  $3.2 \pm 0.6$ ). The measures of fluid cognition, i.e. Raven score (fluid intelligence) and backward digit span (working memory), had a negligible effect on the outcome variable and were not included in the model.

In addition to the fixed effects discussed above, the model contains random intercepts for participants as well as for items (see Table 7.1(b)). The random intercepts for participants model whatever systematic by-participant differences in cognate guessing success that are not entirely covered by the fixed effects in the model. These by-participant differences are assumed to be drawn from a normal distribution with  $\mu$  (i.e. mean) = 0 and an estimated standard deviation of  $\hat{\sigma} = 0.68$  (in log-odds). Similarly, the random intercepts for items model by-item differences in baseline 'translatability'. These are also assumed to be drawn from a normal distribution with  $\mu = 0$  but with  $\hat{\sigma} = 2.0$  log-odds. At this stage, it bears repeating that the full-fledged item analysis is the subject matter of Part III; for now, let it merely be noted that items

differ in their translatability without going into any detail as to why that is the case.

Additionally, it is conceivable that target words differ in their susceptibility to language skills or cognitive abilities as discussed in Section 4.3.1. The corresponding by-item adjustments to the slopes of the covariates can be included in the model. Model comparisons suggested that by-item adjustments for two covariate effects were substantial enough to warrant their inclusion in the final model. First, the random slope for English proficiency by item indicates that higher English proficiency levels are of greater use when translating some items than others. Second, the random slope for WST score by item suggests that high crystallised intelligence levels stand the participants in better stead when decoding some items compared to others. By-item adjustments to the effect of the number of foreign languages in the participants' repertoires or to the effect of correct profile word translation did not improve the model fit and were left out. Such interactions between item- and participant-related characteristics will likewise further be explored in Part III.<sup>18</sup>

## 7.2 Spoken items

6,839 responses to written target items were included in the analyses. All continuous variables were centred at their sample means. The exploratory stage of the analysis indicated that the effects of the covariates could satisfactorily be described using linear terms. The subsequent analyses are therefore carried out using the generalised linear mixed model. The effects of sex, profile word translation and number of foreign languages were small and non-significant and were not included in the final model.

The model's fixed effects are presented in Table 7.2(a) on page 92. Their partial effects are plotted in Figure 7.2 on page 93. The effects of the continuous predictors differ markedly from those for the written items in a few respects. First, the number of languages in the participants' repertoire did not contribute to the model fit and was left out. Conversely, Raven score and backward digit span are now significant predictors. Indeed, Raven score, which is positively associated with cognate guessing

---

<sup>18</sup>`lme4`'s `lmer()` function also computes  $\hat{\rho}$  parameters that model the intercorrelations among the various random effects. These correlation parameters are not of further relevance for the present study. Since they are awkward to report and discuss succinctly, I chose not to present them here in the interest of clarity.

success, has the largest effect size of all predictors involved (ES:  $1.9 \pm 0.4$ ). Note that the effect of backward digit span is a negative one, i.e. higher performance on the backward digit span task is associated with decreased cognate guessing success. Its effect size is comparatively small, however (ES:  $-0.86 \pm 0.41$ ). Furthermore, performance on the WST seems to be positively associated with cognate guessing accuracy, but its effect size (ES:  $1.5 \pm 0.5$ ) is much lower than the case of the written items (ES:  $3.2 \pm 0.6$ ). The effect sizes of English proficiency, however, are very similar between the two modalities (ES:  $1.2 \pm 0.4$  and  $1.5 \pm 0.4$ , i.e. within one standard error of one another).

Table 7.2(b) summarises the model's random effect parameters. In addition to random intercepts for participants ( $\hat{\sigma} = 0.68$ ) and items ( $\hat{\sigma} = 2.5$ ), the model features by-item random slopes for the effects of English proficiency, WST score and Raven score. To repeat, the between-item differences and the interactions between item- and participant-related variables are the subject of Part III.

### 7.3 Variable-by-modality interactions

As discussed above, a comparison of the parameter estimates of the fixed effects in Tables 7.1(a) and 7.2(a) suggests that the effects of several covariates may differ between the two modalities. Specifically, the effects of the number of foreign languages known, WST score, Raven score and backward digit span may be different in the two modalities. The effect of English proficiency, on the other hand, seems to be highly similar across modalities. It is difficult to compare the results of analyses carried out on two subsets of the data directly, however: if a variable has a significant effect in one subset but not in the other, these two effects need not be significantly different *from each other* (Gelman and Stern, 2006; Nieuwenhuis et al., 2011). In order to assess whether any variables have significantly different effects depending on target word modality, I modelled both subsets jointly and assessed whether these variables interact significantly with stimulus modality. Stimulus modality was added to the model both as a fixed main effect and as a by-participant random effect.

The resultant model is not reported in full here as it does not offer any new insights not provided by the modality-specific models reported earlier; Table 7.3 on page 94 presents only the size of the variable-

**Table 7.2:** Generalised (logistic) mixed-effect model modelling translation accuracy on spoken target words in function of participant-related linguistic and cognitive predictors. Panel (a): Fixed effects, their two-tailed significance and their effect sizes. Panel (b): Modelled standard deviation of the random effects ( $\hat{\sigma}$ ). All continuous variables were centred at their sample means. Parameters and effect sizes are expressed in log-odds and are reported to two significant digits.

(a) Fixed effects

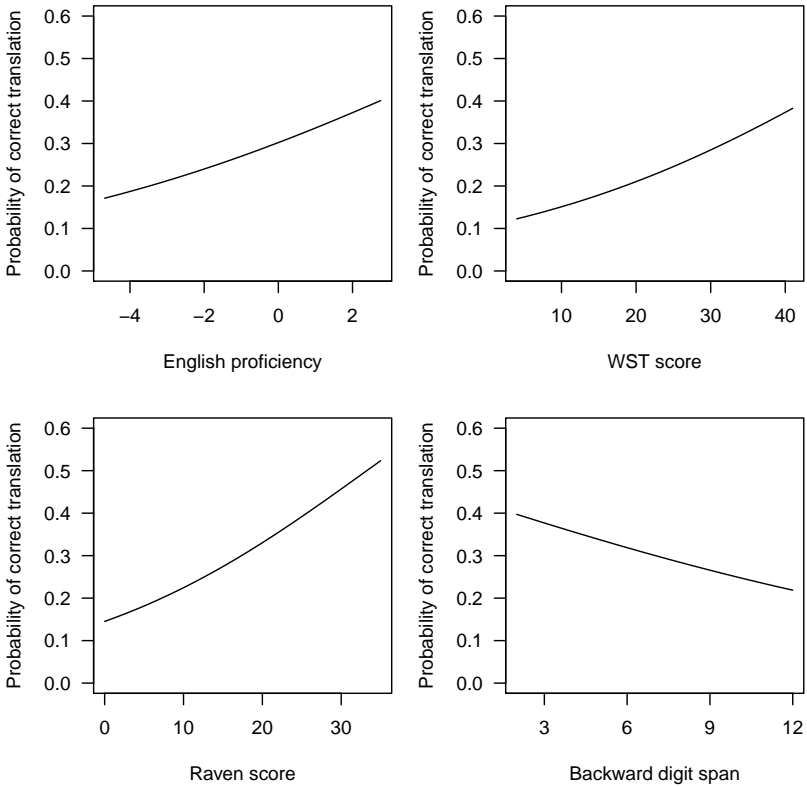
	Estimate $\pm$ SE	<i>p</i>	Effect size $\pm$ SE <sup>a</sup>
Intercept <sup>b</sup>	-1.1 $\pm$ 0.4	0.006	
English proficiency	0.16 $\pm$ 0.06	0.006	1.2 $\pm$ 0.4
WST score	0.040 $\pm$ 0.015	0.006	1.5 $\pm$ 0.5
Raven score	0.053 $\pm$ 0.011	<0.001	1.9 $\pm$ 0.4
Backward digit span	-0.086 $\pm$ 0.041	0.035	-0.86 $\pm$ 0.41

<sup>a</sup> Effect sizes were computed as the largest absolute difference in the outcome variable (in log-odds) when the predictor variable is allowed to vary along its range. See Table 7.1(a) for an example.

<sup>b</sup> The intercept represents the predicted probability of a correct target word translation by a randomly chosen participant whose covariate values correspond to the sample means.

(b) Random effects

	$\hat{\sigma}$
Random intercept by participant	0.68
Random intercept by item	2.5
Random slope for English proficiency by item	0.15
Random slope for WST score by item	0.056
Random slope for Raven score by item	0.029



**Figure 7.2:** Partial fixed effects of the GLMM modelling translation accuracy on spoken target items in terms of participant-related linguistic and cognitive predictors. Continuous variables not included in a given plot were fixed at their medians.

**Table 7.3:** Variable-by-modality interactions. Differences represent the change in the parameter estimates for spoken items relative to written items when responses in both modalities are modelled jointly. Parameters estimates are in log-odds and are reported up to two significant digits.

	Difference $\pm$ SE	<i>p</i>
Intercept	-0.17 $\pm$ 0.50	0.732
$\geq 1$ correct profile word translations	-0.38 $\pm$ 0.21	0.071
Male participant	-0.19 $\pm$ 0.16	0.233
Number of foreign languages	-0.14 $\pm$ 0.09	0.121
English proficiency	0.015 $\pm$ 0.076	0.839
WST score	-0.043 $\pm$ 0.020	0.030
Raven score	0.045 $\pm$ 0.013	<0.001
Backward digit span	-0.091 $\pm$ 0.049	0.061

by-modality interactions as well as their two-tailed significance. I am interested primarily in the by-modality interactions involving the continuous variables, which can be described as follows. The impact of English proficiency is, for all intents and purposes, identical in the two modalities. The roles of crystallised intelligence (WST score) and fluid intelligence (Raven score), on the other hand, are most likely different between the two modalities: fluid intelligence is more important in the spoken modality than in the written modality, whereas the contribution of crystallised intelligence is lower in the spoken than in the written modality. The effect of the number of foreign languages known may likewise be weaker in the spoken modality, and the effect of the working memory capacity measure (backward digit span) may differ according to modality, too, but the by-modality interactions of both variables are not significant at the conventional 0.05 threshold.

# Chapter 8

## Discussion

### 8.1 Age trends

The first goal of this thesis was to track the lifespan development of  $Lx$  cognate guessing skills. Previous studies had suggested that the ability to correctly guess the meaning of words in an unknown but related language improves throughout childhood and adolescence, but it was not known how this ability develops throughout the adult lifespan (Section 2.4). The present study approached this question systematically by administering a written and a spoken cognate guessing task to a sample of multilingual participants aged 10 to 86 years.

On the one hand, the results of this study fit in with previous findings in that it found an increase in the ability to correctly translate both written and spoken cognates from an unknown related language throughout childhood and adolescence. At the same time, this study shows how  $Lx$  cognate guessing skills continue to develop past young adulthood: for written target words, cognate guessing success is at least stable from about age 20 onwards and may even show continued gradual increase throughout the adult lifespan (see Figure 6.2(a) on page 82). For spoken stimuli, cognate guessing success appears to be fairly stable from age 20 to about age 50, at which point performance starts to drop (see Figure 6.2(b)).

## 8.2 Linguistic and cognitive predictors

Welford (1958, p. 14; cited in Salthouse, 2006, p. 276) suggested that the extent to which age-related trends in complex task performance reflect the developmental trajectories of fluid and crystallised intelligence measures depends on the demands that the task in question places on these facets of intelligence—a suggestion which Salthouse (2006) describes as “probably . . . readily accepted by most contemporary researchers” (p. 276). My second goal, therefore, was to investigate to what extent the age trends in cognate guessing skills could be explained by the impact of crystallised and fluid resources on cognate guessing skills and how the age-by-modality interaction described in the previous section could be accounted for.

The variables considered were (a) the number of foreign languages in the participants’ repertoires, (b) English proficiency, (c) crystallised intelligence, (d) fluid intelligence and (e) working memory. (d) and (e) are uncontroversially considered fluid resources, whereas (c) is a crystallised resource. (a) and (b) can be considered crystallised resources as well since they are to a substantial extent the product of learning and experience. Since these variables are expressed on different scales, their effects are best compared with reference to the effect sizes provided in Tables 7.1 and 7.2 on pages 87 and 92, respectively.

When doing so, it becomes clear that the number of foreign languages in the participants’ repertoires is at best a modest predictor of their cognate guessing ability. In the written modality, its effect is merely  $0.62 \pm 0.29$  log-odds, which is markedly lower than that of English proficiency and WST score, whereas its effect in the spoken modality is negligible. In fact, the effect of this variable may not differ by modality as suggested by the non-significance of the interaction effect presented in Table 7.3. These results indicate that the sheer number of languages in the participants’ repertoires does not strongly affect their cognate guessing ability. Several studies, however, presented evidence that one specific form of multilingualism can yield such an advantage, namely proficient multilingualism in language varieties closely related to one another and to the *Lx* (Section 2.1.2). In the present study, only two participants had some minimal knowledge of a Germanic foreign language beside English, viz. Dutch, making it difficult to gauge the influence of this factor in a quantitative analysis.

The effect of proficiency in the Germanic foreign language English, by contrast, could be assessed quantitatively. English proficiency is a respectable predictor of  $Lx$  cognate guessing success with effect sizes of  $1.5 \pm 0.4$  and  $1.2 \pm 0.4$  log-odds in the written and spoken modalities, respectively. This result meshes with findings that ‘closely related multilingualism’ is advantageous in receptive multilingualism: as a Germanic language that is closely related to both the  $Lx$  in question (Swedish) and the participants’ native languages (Swiss German and standard German), English can provide useful transfer bases and may speculatively serve as a vehicle for abductive processes as discussed in Section 2.1.2. English proficiency shows a marked age trend in the sample in that it rises steeply up to about age 30 and declines from that point onwards (see Figure 5.2 on page 72). Its role, however, is similar in both modalities (see also Table 7.3). This means that it cannot be the main underlying cause for the age-by-modality interaction in  $Lx$  cognate guessing skills. But since English proficiency, like the other four variables considered, shows an age-related increase up to about age 30, it is likely that it contributes to the concomitant increase in cognate guessing success in both modalities.

To explain the differential age trends in cognate guessing in terms of linguistic and cognitive factors, the other variables thus need to be considered. The stability, or indeed improvement, of written cognate guessing success throughout the adult lifespan suggests a dependence on mainly crystallised resources, whereas the age-related demise of spoken cognate guessing success indicates a stronger reliance on fluid resources (as per Welford, 1958). Indeed, the most important predictor of written cognate guessing success is the crystallised intelligence measure, L1 vocabulary knowledge, with an effect size of  $3.2 \pm 0.6$  log-odds. This finding is consistent with suggestions that a well-developed L1 vocabulary is conducive to one’s ability to understand closely related languages (see Section 3.1.3). Simultaneously, it may be interpreted in terms of Berthele’s (2011) suggestion that  $Lx$  cognate guessing draws partly on abductive processes (see Section 2.1.2): participants with a large L1 vocabulary and, more generally, broad experience in coping with linguistic variation of sundry kinds (regional, social, stylistic etc.) may be in a better position to speculate about plausible inter-varietal form correspondences. In the spoken modality, crystallised intelligence is still a respectable predictor of  $Lx$  cognate guessing, but with an effect size of  $1.5 \pm 0.5$  log-odds, its impact is appreciably weaker (see also Table 7.3).

Reliance on crystallised intelligence can consequently provide stability in written  $Lx$  cognate guessing throughout the adult lifespan, but less so in spoken  $Lx$  cognate guessing. I will discuss why such by-modality differences might arise below.

As for the fluid resources, it was found that the effects of fluid intelligence and working memory were negligible in the written modality. In the spoken modality, by contrast, fluid intelligence turned out to be the most important predictor of all with an effect size of  $1.9 \pm 0.4$  log-odds. Decreases in fluid intelligence from about age 30 onwards (see Figure 5.2 on page 72) may therefore have contributed to the age-related decline in spoken  $Lx$  cognate guessing skills whilst not affecting written  $Lx$  cognate guessing. I posited that a contribution of fluid intelligence to  $Lx$  cognate guessing could be expected on the grounds that fluid intelligence represents a person's skills to deal flexibly with new information and to solve novel problems creatively. Translating isolated Swedish words into German can safely be assumed to have been a task not previously encountered by our participants. (By contrast, engaging in receptive multilingualism more generally, including between varieties belonging to the same language, cannot.) Moreover, abductive reasoning may be necessary to cope with obfuscated formal resemblances between the  $L1$ ,  $L2$ , ...,  $L_n$  and the  $Lx$ , and such abductive reasoning may draw on the participants' fluid intelligence. I will turn to the question of why fluid intelligence has a differential effect according to stimulus modality below.

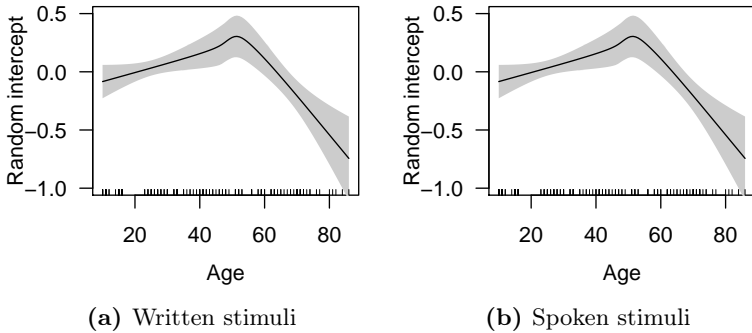
Curiously and contrary to expectations, working memory was actually *negatively* associated with spoken cognate guessing success when considered jointly with other linguistic and cognitive predictors in a multivariate model (see Table 7.2). In the written modality, by contrast, the effect of working memory was negligible, and its effect was not significantly different between the two modalities (Table 7.3). At present, I am hesitant to venture a theoretical post-hoc explanation for the negative effect found for spoken target words given that it clashes with my predictions and given that the effect size is comparatively small ( $-0.86 \pm 0.41$  log-odds). To put it in Bayesian terms, the prior probability of a negative effect was all but zero and a small negative effect in the data is not sufficient to lead me to believe that an actual negative effect exists. I therefore prefer to offer this finding as a spur for a future study, which could measure working memory capacity more stringently than I

did, for instance by means of a latent variable analysis (see the discussion on page 53).<sup>19</sup>

Having established that Lx cognate guessing skills follow a different developmental trajectory depending on stimulus modality and that a differential reliance on fluid and crystallised intelligence in particular underlies this difference to a substantial degree, I now turn to the question of why fluid and crystallised resources should impact cognate guessing in the two modalities differentially to begin with. Speculatively, it may be more cognitively challenging to compare phones and phonemes across languages than letters and graphemes. An alternative explanation is that it may be the time pressure associated with auditory stimulus presentation that causes the difference. Spoken items were presented just once, whereas written items remained on-screen until the participants entered their translations. Spoken items thus required above all the quick application of cognitive flexibility, whereas in the written modality, cognitive speed was a lesser issue and participants had more time to engage in “linguistisches Probabilitätskalkül” (Berthele, 2008, p. 92). Thus, differences in time pressure may, on the one hand, account for the presence of a fluid intelligence effect in the spoken modality and the absence of such an effect in the written modality and, on the other hand, for the greater importance of crystallised resources, which are assumed to underlie such “linguistisches Probabilitätskalkül”, in the written modality. A similar point about full text comprehension in a related foreign language was made by Ringbom (1992, p. 94), who argued that the absence of time constraints in reading enables readers to draw on knowledge resources more than in listening. The implication of this suggestion is that repeated aural presentation might conceivably have diminished the differential effect of modality since it would have lowered the time pressure associated with the spoken modality. In order to guarantee a minimal ecological validity of the task, however, the stimuli were presented only once—just as in oral communication, where words and sentences are uttered once and not repeated unless there is a pragmatic reason to do so.

---

<sup>19</sup>For such purposes, it bears pointing out that the effect of working memory is, in fact, positive when it is analysed independently of other participant-related variables, presumably due to its inter-correlations with the other predictors (see Figure 5.3 on page 76).



**Figure 8.1:** Residual age effects in the random intercepts.

### 8.3 Residual age trends

The discussion of the results thus far raises the question of how adequately the age effects in  $Lx$  cognate guessing skills can be explained by the participant-related linguistic and cognitive factors considered. The by-participant random intercepts in the models summarised in Tables 7.1 and 7.2 indicates that these models did not fully capture inter-individual variation in  $Lx$  cognate guessing success. Capturing all inter-individual variation is obviously an unattainable ideal, but since my goal was to model the age trends found in Chapter 6 in terms of participant-related linguistic and cognitive factors, this residual inter-individual variation does spur the question whether I fully succeeded in modelling these age trends.

Residual inter-individual variation should largely have been captured by the by-participant random intercepts in the GLMMs. The amount of variance in these random effects that can be accounted for by the participants' age can be gauged using GAMs. I therefore fitted separate (Gaussian) GAMs on the by-participant random intercepts in the models presented in Tables 7.1 and 7.2. Age was the sole predictor in these GAMs, and its effect was modelled using an adaptive smooth (see Note 17 on page 79). The two GAMs modelling residual age trends in the random intercepts are presented in Figure 8.1.

Clearly, I was not wholly successful in modelling the age trends in cognate guessing success in terms of participant-related linguistic and

cognitive predictors. As can be seen in Figure 8.1(a), the fixed effects in the GLMM for written stimuli jointly underestimate the increases in cognate guessing task performance throughout adulthood. More specifically, they seem to overestimate the performance of the participants in the 20-to-30 age bracket relative to that of the younger and older participants, hence the negative correction imposed by the random intercepts. However, this age effect can account for merely 4.2% of the variance in the random intercepts and is not significant ( $F = 1.2$ , est.  $df = 3.1$ ,  $n = 163$ ,  $p = 0.31$ ). By contrast, the fixed effects of the participant-related linguistic and cognitive predictors simultaneously seem to slightly underestimate the age-related increase in spoken cognate guessing task performance up to about age 50 and (more markedly) the decrease from that age onwards. This age effect is significant ( $F = 6.9$ , est.  $df = 2.9$ ,  $n = 163$ ,  $p < 0.001$ ) and accounts for 13.4% of the variance in the random intercepts.

In sum, age trends can still account for a relatively small proportion of inter-individual differences that are not captured by linear trends of linguistic and cognitive variables. This is particularly the case for performance on the auditory task. It is plausible that more fine-grained measures of the constructs that were considered would have explained more inter-individual variance or that these residual age trends are linked to linguistic and cognitive factors that were not considered in these analyses, e.g. cognitive control (see Section 3.3). One such factor that may go some way in explaining the residual age trends in auditory cognate guessing is the participants' hearing acuity. While all participants reported normal or corrected-to-normal hearing, hearing acuity decreases in older age may go undetected in self-reports (see Gordon-Salant, 2005, pp. 17–18). That said, the auditory stimuli were presented over headphones and participants could notify the experimenters of any difficulties after a 5-stimulus training run. The possible impact of hearing acuity decreases on the findings presented here should therefore not be overstated.

## 8.4 Outlook

The present part of this dissertation focussed exclusively on participant-related characteristics. However, research on receptive multilingualism is also interested in the impact of stimulus-related characteristics on cognate guessing accuracy. With its fairly large dataset of 163 participants

guessing the meaning of 90 *Lx* cognates each, this project can contribute to a better understanding of how stimulus-related characteristics affect cognate guessing accuracy in multilinguals.

A unique aspect of the present project is that it covers a wide range of participants in terms of age, language skills and cognitive abilities. It is conceivable that the effects of certain stimulus-related properties vary in function of these participant-related characteristics or, equivalently, that the effects of participant-related characteristics vary in function of certain stimulus-related properties. The by-item random slopes in Tables 7.1 and 7.2 indicate that the effects of participant-related variables do indeed vary from item to item. In Part III, I investigate which item-related variables affect *Lx* cognate guessing accuracy in the written and spoken modalities and whether these variables interact with the participant-related characteristics discussed in the present part.

## Part III

# Integrating item-related characteristics



## Chapter 9

# Item-related determinants of cognate guessing

In the previous part, I focussed on the effects of participant-related variables on cognate guessing accuracy. However, much research on receptive multilingualism and cognate guessing is also concerned with identifying the impact of stimulus-related variables, such as word frequency or the degree of overlap to a known word, on cognate guessing. An interesting possibility to consider is that the effects of these stimulus-related variables may vary systematically in function of some of the participant-related variables investigated in the previous part. If such item–participant interactions do indeed exist, this in turn gives rise to the possibility that the effects of stimulus-related variables vary across the lifespan, too. This puts the exploration of item–participant interactions firmly within the scope of the present study, and it is this exploration that I will undertake in this and the following chapter.

Previous research on receptive multilingualism has identified a great many item-related variables that may influence cognate guessing accuracy. This research is summarised in Section 9.1. Since it is impractical to investigate the interactions between all of these variables and the participant-related variables, the first priority is to select from this multitude of item-related variables a handful of predictors that most parsimoniously accounts for the between-item variance in cognate guessing accuracy. This is the object of Sections 9.2 to 9.4. The results of

this investigation are interesting in their own right and are discussed in Section 9.5. Then, in Chapter 10, I investigate how this handful of predictors interacts with age and other participant-related variables.

## 9.1 Previous findings

### 9.1.1 Formal distance between cognates

As discussed in Section 2.1.1 on page 18, language transfer is particularly likely to take place when the language user *perceives* similarities between the source language and the target language. While perceived similarity does not equate to (objective) formal similarity, the two concepts indisputably overlap to a large extent so that formal similarity can be used as a first rough approximation of perceived similarity. A natural assumption when modelling  $Lx$  text or word comprehension is consequently that the degree of formal overlap between the  $Lx$  text or word and its  $L1$  ( $L2, \dots, Ln$ ) counterpart is a highly important factor. When the two show an almost perfect formal overlap, comprehension should be at ceiling. Conversely, when little overlap exists, readers or listeners need to fill in the gaps themselves, which increases the likelihood that comprehension will drop. In what follows, I briefly review how the formal distance between  $Lx$  and  $L1$  ( $L2, \dots, Ln$ ) stimuli is typically quantified and how well such measures can predict  $Lx$  stimulus comprehension.

#### Quantification

The formal distance between an  $Lx$  word and its cognate in a known language is often measured by means of the Levenshtein (1966) algorithm. This algorithm was initially adopted from dialectometry (e.g. Heeringa, 2004; Kessler, 1995) by Van Bezooijen and Gooskens (2005a,b) and is used to compute the smallest number of deletions, insertions and substitutions necessary to transform one string into another. The Levenshtein distance is defined as the total operation cost, i.e. the number of necessary deletions, insertions and substitutions. For the computation of phonetic Levenshtein distances, phonetic transcriptions are used as input strings; for orthographic distances, letter strings are used. An example is given in Figure 9.1 on the next page, which shows how the Swedish phonetic string [ensamhert] (*ensamhet* ‘lonesomeness’) can be transformed into

1	2	3	4	5	6	7	8	9	10	
ε		n	s	a	m	h	e:		t	
a	I	n	z	a	m	k	a	I	t	
S	I		S			S	S	I		= 6

**Figure 9.1:** Example of a Levenshtein distance computation. The Swedish phonetic string [ɛnsamhɛ:t] (*ensamhet* ‘loneliness’) is transformed into German [amzɑmkɑ:t] (*Einsamkeit*) with a minimal conversion cost of four substitutions (S) and two insertions (I).

German [amzɑmkɑ:t] (*Einsamkeit*) with a minimal total transformation cost of 6.

Raw Levenshtein distances are not very useful when modelling Lx comprehension: long Levenshtein alignments, by their very nature, tend to be associated with high Levenshtein distance values, but long Lx strings also contain more phonetic or orthographic information that can be of help to readers or listeners. The pair [ɛnsamhɛ:t]–[amzɑmkɑ:t], even though associated with a raw Levenshtein distance of 6, is arguably more transparent than, for instance, the pair [gɔ:]–[gɛ:ɔn] (*gå–gehen* ‘to go’), which has a raw Levenshtein distance of 3. Raw Levenshtein distances are therefore typically length-normalised in order to take this bias into account. This can straightforwardly be accomplished by dividing the raw distance score by the length of the longer string involved. For the pair [ɛnsamhɛ:t]–[amzɑmkɑ:t], length-normalising the raw Levenshtein distance in this way yields a normalised score of 0.60; for the pair [gɔ:]–[gɛ:ɔn], the normalised value is 0.75. However, Heeringa (2004, pp. 130–132) argued that Levenshtein distances should be normalised by dividing them by the length of the least-cost alignment rather than by string length. In many cases, this yields identical normalised scores, but consider, for example, the Swedish–German cognate pair [sykɛl]–[tsyklʊs] (*cykel–Zyklus* ‘cycle’) in Figure 9.2 on the following page. In order to transform one cognate into the other, a minimum of four operations are needed, but this can be accomplished in either a 7-slot or an 8-slot alignment. Normalising the raw Levenshtein distance of 4 by string

1	2	3	4	5	6	7	
	s	ʏ	k	ɛ	l		
t	s	ʏ	k	l	ʊ	s	
I				S	S	I	= 4
(a) 7-slot alignment							
1	2	3	4	5	6	7	8
	s	ʏ	k	ɛ	l		
t	s	ʏ	k		l	ʊ	s
I				D		I	I
= 4							
(b) 8-slot alignment							

**Figure 9.2:** The Swedish phonetic string [svkɛl] (*cykel* ‘(bi)cycle’) can be transformed into German [tsvklʊs] (*Zyklus* ‘cycle’) with a minimum of four operations. Both a 7-slot or an 8-slot alignment can be used, the 8-slot alignment arguably making more sense for linguistic purposes.

length yields a score of 0.57, but normalising by alignment length yields normalised scores of 0.57 or 0.50 depending on which alignment is chosen. As Heeringa (2004) argued, the 8-slot alignment is more sensible from a psycholinguistic point of view as it contains more phone matches than the 7-slot alignments: “We suppose that in perception people will try to match the common sounds in two different pronunciations, so we prefer the longer alignments” (p. 131). Normalisation by the length of the longest least-cost alignment is therefore typically preferred in studies on receptive multilingualism.

In its crudest form, the Levenshtein algorithm is insensitive to linguistic correspondences, but it can be refined in order to mirror more closely how listeners perceive linguistic stimuli. On a general level, for instance, only vowel–vowel and consonant–consonant matches can be allowed (e.g.

Beijering et al., 2008; Gooskens et al., 2008), but more subtle differences can be modelled as well by assigning different operation weights depending on the phonetic or phonological correspondences between the aligned phones (e.g. Beijering et al., 2008; Gooskens, 2007b; Gooskens et al., 2008). Orthographic Levenshtein distances are rarely tweaked in a similar fashion (but see Van Bezooijen and Gooskens, 2005a), but if the languages in question differentiate consistently between vowel and consonant graphemes, forced vowel–vowel and consonant–consonant matches can straightforwardly be implemented in the written modality as well.

### Predictive power

**Spoken modality** A handful of previous studies investigated the link between  $Lx$  stimulus comprehension and phonetic Levenshtein distances between the  $Lx$  items and their translation equivalents in known languages (usually the L1). Gooskens (2007b), for instance, reported an overall correlation of  $r = -0.64$  between spoken text comprehension and aggregate phonetic Levenshtein distances between Scandinavian language varieties as well as between the West-Germanic languages Afrikaans, Dutch and Frisian. Thus, larger phonetic distances between the  $Lx$  texts and their L1 translation equivalents were associated with poorer  $Lx$  text comprehension. When only the Scandinavian varieties were considered, the correlation was stronger at  $r = -0.80$ , which is comparable in size to the correlation of  $r = -0.86$  found in a similar study by Beijering et al. (2008, see also Gooskens et al., 2008) that investigated the comprehension of a spoken text in several Scandinavian language varieties by speakers of standard Danish.

Gooskens (2007b) and Beijering et al. (2008) correlated phonetic Levenshtein distances with text comprehension rather than with the comprehension of individual words, making it difficult to partial out the influence of context, prosody and syntax. Kürschner et al. (2008) therefore correlated the comprehension of 347 spoken Swedish words presented in isolation by 42 Danish subjects (aged 16–19 years) with the phonetic Levenshtein distances between these words and their Danish cognates. The correlation found ( $r = -0.27$ ) was markedly weaker than the correlations involving text comprehension discussed above. Kürschner et al. speculated that word-specific idiosyncrasies may affect the comprehension of single words, which in turn may result in a weaker

correlation coefficient. In the aggregate, i.e. on the level of text comprehension, the effects of such idiosyncrasies may average out, potentially resulting in stronger correlations. In a similar study, however, Doetjes and Gooskens (2009) reported a somewhat stronger correlation between phonetic Levenshtein distances and the comprehension of 76 isolated Swedish words with Danish cognates by 54 Danish subjects aged 16 to 19 years ( $r = -0.54$ ). Gooskens et al. (2011) reported a correlation of similar strength ( $r = -0.61$ ) in a study investigating the comprehension of 369 Low German words (a mixture of cognates and non-cognates) by 124 Dutch 15- to 18-year-olds.<sup>20</sup> In addition, Van Bezooijen and Gooskens (2005a) had earlier reported an even stronger correlation of  $-0.74$  between the comprehension of 32 Afrikaans and Frisian words by 67 Dutch school pupils (mean age: 16.3 years) and the phonetic Levenshtein distances between these words and their Dutch cognates.

As I pointed out earlier (Section 2.1.2 on page 19), L1 knowledge of a language related to the  $Lx$  does not preclude readers and listeners from drawing on their L2, ...,  $L_n$  knowledge, too. Reliance on L2, ...,  $L_n$  knowledge can reflect itself in the correlations between  $Lx$  word comprehension and their Levenshtein distance towards the L2, ...,  $L_n$  as illustrated by a study by Berthele (2011). He asked 163 Swiss German subjects aged 13–35 years to translate 28 Danish and Swedish verbs with related translation equivalents in German or English or both. While he did not find a significant correlation between  $Lx$  word comprehension and phonetic Levenshtein distances between the Scandinavian targets and their standard German translation equivalents ( $r = 0.00$ ), the correlation with the Levenshtein distances to the  $Lx$  targets' English translation equivalents was in the expected direction ( $r = -0.35$ ). English was the participants' L2 or L3, and these results therefore suggest that the participants may be sensitive to  $Lx$ –L2, ...,  $L_n$  correspondences as well.

**Written modality** Orthographic Levenshtein distances have been used in order to account for specific patterns in written  $Lx$  text comprehension. Afrikaans and Frisian are both West-Germanic languages like Dutch, yet Van Bezooijen and Gooskens (2005a,b) found that speakers of Dutch are able to understand written Dutch texts translated into

---

<sup>20</sup>Gooskens et al. (2011) also investigated the comprehension of standard (High) German by Dutch pupils, but due to schooling and intensive exposure, High German cannot be considered an  $Lx$  to these participants.

Afrikaans better than the same texts translated into Frisian. In line with this finding, the aggregate Levenshtein distances that Van Bezooijen and Gooskens (2005a,b) computed turned out to be higher between the original Dutch text and its Frisian translation than between the Dutch text and its Afrikaans translation. Additionally, Gooskens and Van Bezooijen (2006) found that Afrikaans–Dutch written text comprehension is asymmetric: Dutch participants performed better on the Afrikaans task than Afrikaans-speaking participants did on the Dutch task. Again, Levenshtein distances turned out to be marked by a similar skew.<sup>21</sup>

Taken together, the findings from these three studies suggest that orthographic Levenshtein distances may be able to account for some proportion of the variance in  $Lx$  stimulus comprehension. Unfortunately for the present purposes, their comparisons of whole texts do not yield a particularly clear estimate of precisely how useful such distance computations are for predicting the comprehension of isolated  $Lx$  words. More relevant from this perspective is a study by Berthele and Lambelet (2009), who investigated how well 140 French- and Italian-speaking Swiss students understood 29 isolated Romansh and Romanian words. They found that comprehension was correlated with the crude orthographic Levenshtein distance between the  $Lx$  word in question and its French or Italian cognate ( $r = -0.32$ ; the authors used whichever of the two distances was the lower one). In a similar task involving 28 Danish and Swedish isolated words and 163 Swiss German participants, Berthele (2011) did not find any association between orthographic Levenshtein distances between the Scandinavian words and their written (standard) German cognates ( $r = 0.14$ ), whereas such a correlation was present when the orthographic distances with respect to English, the participants' L2 or L3, were considered ( $r = -0.42$ ; for similar findings for phonetic Levenshtein distances by Berthele, 2011, refer to page 110). Again,

---

<sup>21</sup>The Levenshtein distance for any given word pair is symmetric, i.e. it takes as many operations to transform any given string into any other given string as vice versa. Still, it is possible for the Levenshtein distance between two texts to be asymmetric. For instance, Gooskens and Van Bezooijen's (2006) Dutch text contained the word *vaak* 'often', which was aligned with the Afrikaans non-cognate *dikwels*. Speakers of Afrikaans may not be able to arrive at the meaning of *vaak*. Speakers of Dutch, however, will most likely be able to derive the meaning of *dikwels* since a synonymous cognate exists in Dutch, viz. *dikwyls*. In this case, the Levenshtein distance from L1 Afrikaans (*dikwels*) to  $Lx$  Dutch (*vaak*; 7 operations, length-normalised: 0.875) is higher than the one from L1 Dutch (*dikwyls*) to  $Lx$  Afrikaans (*dikwels*; 2 operations; length-normalised: 0.25).

these results suggest that participants draw on their foreign-language knowledge in  $Lx$  cognate guessing, even when the  $Lx$  is related to their L1.

### 9.1.2 The importance of consonants

By default, the Levenshtein algorithm weights all operations equally, i.e. a vowel–vowel substitution has the same operation weight as a consonant–consonant substitution. However, a few studies suggest that  $Lx$  comprehension is more severely affected by consonantal differences than by vocalic differences between the L1 ( $L2, \dots, L_n$ ) and the  $Lx$ . Gooskens et al. (2008) computed Levenshtein distances between a spoken standard Danish text and its renderings in 17 other Scandinavian language varieties. They then computed how much of these distances was due to vowel insertions, deletions and substitutions (vocalic Levenshtein distances) and how much was due to consonant operations (consonantal Levenshtein distances). Thus, in the *ensamhet–Einsamkeit* example in Figure 9.1 on page 107, the overall Levenshtein distance is  $6 \div 10 = 0.60$ , the vocalic Levenshtein distance  $4 \div 10 = 0.40$  (substitutions of [ɛ] and [e:] and two [i] insertions) and the consonantal Levenshtein distance  $2 \div 10 = 0.20$  (substitutions of [s] and [h]). The authors found that the comprehension of the 17 Scandinavian language varieties by 351 speakers of standard Danish correlated more strongly with consonantal than with vocalic Levenshtein distances (consonants:  $r = -0.74$ , vowels:  $r = -0.29$ ), although the correlation with the overall phonetic Levenshtein distances was even stronger ( $r = -0.86$ ).

Similarly, Berthele (2011) found consonantal differences vis-à-vis German and English to be more detrimental to how well Swiss Germans understood isolated written and spoken Danish and Swedish words compared to vocalic contrasts. These results corroborate findings by Möller (2011; see also Möller and Zeevaert, 2010), who reported that German-speaking students show greater tolerance towards vocalic differences between  $Lx$  stimuli and their L1 cognates than towards consonantal differences when trying to come up with German cognates of individually presented written words in other Germanic languages.

The greater flexibility vis-à-vis vocalic discrepancies in receptive multilingualism studies can be linked to similar findings stemming from L1 tasks. Of particular interest is a series of experiments conducted by Van Ooijen (1996) with English-speaking subjects. Participants were

presented with auditory nonsense strings, e.g. [ˈfɒŋgə], that could be transformed into existing English words both by changing one vowel segment ([ˈfɪŋgə] *finger*) or by changing one consonant segment ([ˈlɒŋgə] *longer*). When asked to transform the nonsense strings into words by changing vowel phonemes, participants responded faster and more correctly than when asked to produce words by changing consonant phonemes. Additionally, vocalic substitutions were associated with shorter response latencies in the free-choice condition as well. Similar results are reported by Cutler et al. (2000) for Dutch and Spanish participants, by Marks et al. (2002) for monolingual Spanish-speaking participants and for English–Spanish bilinguals and by Cutler and Otake (2002) for Japanese participants. These findings suggest that listeners from different linguistic backgrounds perceive vowel identities to be more mutable than consonant identities even in auditory L1 word recognition. Furthermore, Moates and Marks (2012) showed that vowel mutability generalises to the written modality, at least for English and Spanish speakers.

It is not clear what causes listeners to be more flexible towards vocalic discrepancies than towards consonantal discrepancies. Various hypotheses have been advanced (see Moates and Marks, 2012, and Van Heuven, 2008), but they are difficult to reconcile with the fact that vowel mutability is observed for speakers of languages with substantially different structural properties (see Moates and Marks, 2012). Moreover, differential treatment of vowels and consonants does not seem to be a universal feature of the lexical activation process proper: while some L1 reading experiments with native speakers of English suggest that consonantal information is derived faster and more automatically from grapheme strings than is vocalic information (Berent and Perfetti, 1995; Lee et al., 2001), not all experiments unequivocally support such a hypothesis of consonantal primacy in languages other than English (see Lee et al., 2001, pp. 199–200) or indeed in English (Lukatela and Turvey, 2000; Perry and Ziegler, 2002). Moreover, in two auditory priming experiments with speakers of English and Dutch, Cutler et al. (1999) did not find any evidence suggesting that lexical entries are activated more when primed by non-words with the same consonants but one different vowel phoneme than by non-words with the same vowels but one different consonant phoneme. Therefore, “it is most probably in the decision processes involved in the alteration operation that the

vowel/consonant differences observed with the reconstruction task are located” (Cutler et al., 1999, p. 2055). Given that the *Lx* cognate guessing tasks such as those employed by Berthele (2011), Gooskens et al. (2008) and Möller (2011)—as well as in the present study—tap into a number of decision processes, too,<sup>22</sup> a differential treatment of vowels and consonants can carry over to such tasks. That said, I must leave unaddressed the more basic question as to why there should be such differential treatment of vowels and consonants when decision processes are at play to begin with.

### 9.1.3 The importance of word beginnings

Participants not only seem to let guide their *Lx* cognate guessing attempts more by the consonantal skeleton of the *Lx* stimulus than by its vowels; they may also be more sensitive to word onsets as opposed to rhymes in both visual and auditory cognate guessing tasks (Berthele, 2011; Möller, 2011; Möller and Zeevaert, 2010; see also Müller-Lancé, 2003, and Reinfried, 1998, pp. 39–40). Some parallels can be again drawn to auditory and visual L1 word recognition.

As regards auditory L1 word processing, some models, e.g. COHORT (Marslen-Wilson, 1987; Marslen-Wilson and Welsh, 1978), explicitly weight onsets more heavily than rhymes, whereas others, e.g. TRACE (McClelland and Elman, 1986) and Shortlist (Norris, 1994; Norris and McQueen, 2008), do not. Nooteboom and Van der Vlugt (1988) and Connine et al. (1993) provide evidence that the auditory word recognition architecture does not inherently weight onsets more heavily than rhymes by design, with Connine et al. (1993) concluding that “the lexical item activated in memory is simply the best hypothesis available for the acoustic input” (p. 207). Nevertheless, word onsets tend to contain more contrastive information than do rhymes and be less variant in connected speech than other word parts (see Gow et al., 1996, and Van Heuven,

---

<sup>22</sup>This assumption is hardly open to question. In fact, such decision processes can be quite drawn out in *Lx* decoding tasks: Berthele (2011) reproduces a transcript from a think-aloud session in which the participant took roughly 50 seconds before settling on a final translation. Further note that by using the term ‘decision processes’ I do not mean to imply that the participants necessarily *deliberately* act according to a set of algorithms featuring differential treatment of consonants and vowels—although for some this might well be the case. Rather, what I want to emphasise that this kind of task does not tax the lexical activation architecture in the same way as ordinary lexical decision tasks or priming experiments are assumed to.

2008). Differential treatment of onsets and rhymes may therefore still seemingly emerge as a consequence of different loads of contrastive information in different word parts. In visual L1 word processing, too, word beginnings tend to contain more contrastive information than word endings and seem to be more important in accurate word recognition (Broerse and Zwaan, 1966; Johnson and Eisler, 2012; Scaltritti and Balota, 2013).

### 9.1.4 ‘Exotic’ phones, suprasegmentals and graphemes

Exotic-sounding *Lx*-specific phones, such as the Swedish *sje* sound ([ʃj]), may throw the unaccustomed listener off-balance and thus form a particular hindrance to spoken *Lx* comprehension (Van Heuven, 2008). The effect of such phones was investigated empirically in a Scandinavian context by Kürschner et al. (2008), but was found to be low, as it accounted for only about one percent of the variance in spoken *Lx* stimulus comprehension. Similarly, *Lx*-specific suprasegmentals have occasionally been postulated as being potentially detrimental to spoken *Lx* comprehension (e.g. Bannert, 1981). Two examples are the Danish *stød*, a kind of creaky voice that is phonologically distinctive in minimal pairs such as [hun] (*hun* ‘she’) – [hunʔ] (*hund* ‘dog’), and the Swedish tonal word accents, which are phonologically distinctive in about 350 minimal pairs such as between [ˈandɛn] (*anden* ‘the duck’; said to have ‘accent 1’) and [ˌandɛn] (*anden* ‘the spirit’; said to have ‘accent 2’) (Elert, 1981b, pp. 61–68), most of which would never be confused in actual language use. The impact of these *Lx*-specific suprasegmentals on isolated spoken word comprehension seems to be minimal as well (Gooskens and Kürschner, 2010; Kürschner et al., 2008). Seeing as the effects of such ‘exotic’ features are minimal at best in spoken cognate guessing tasks and given that the Swedish stimulus set used in this study features only four words with the *sje* sound and nine words that are marked by accent 2, I will not consider the effect of these ‘exotic’ features in the analyses.<sup>23</sup>

<sup>23</sup>Note further that according to Elert (1981b), the main function of Swedish tonal word accents is to mark the distinction between compounds (accent 2) and two separate words (accent 1) in connected speech. Thus, to the extent that tonal word accent affects *Lx* stimulus comprehension at all, this is likely to be caused by listeners unfamiliar with this distinction perceiving accent 2 words as two separate words.

To my knowledge, no study has investigated whether ‘exotic’ graphemes negatively affect written *Lx* stimulus comprehension. The only Swedish grapheme that does not occur in German is the *å* character (save in some arcane technical terms such as *Ångström*). The present stimulus set features merely five words containing *å*. I consider this to be too few to be analysed meaningfully in a quantitative model and will not consider this variable in the analyses.

### 9.1.5 Lexical stress differences

On the basis of research on the role of stress in auditory L1 word recognition (see Van Heuven, 1985), Van Heuven (2008) hypothesises that differences in lexical stress positioning between an *Lx* stimulus on the one hand and a known cognate on the other negatively affect stimulus comprehension. According to Cutler (2005), however, the role of lexical stress in L1 word recognition depends on the language studied, with stress being less important in English than in Dutch or German word recognition. Kürschner et al. (2008) investigated the comprehension of isolated Swedish words by speakers of Danish and did not find any convincing evidence in support of the lexical stress hypothesis in *Lx* cognate comprehension, although it should also be pointed out that stress differences between Danish and Swedish cognates are fairly infrequent. Consequently, only 10 out of 347 cognate pairs in Kürschner et al.’s (2008) study were actually characterised by word stress differences, which may well be the primary reason why their findings do not lend support to Van Heuven’s (2008) hypothesis. Similarly, the present stimulus set does not feature a single Swedish–German cognate pair with stress differences, and I will not pursue Van Heuven’s (2008) lexical stress hypothesis any further.

---

Like Gooskens and Kürschner (2010) and Kürschner et al. (2008), I informed my participants that they would hear isolated words, and they were therefore a priori unlikely to give two-word answers.

### 9.1.6 Word frequency, neighbourhood density and word length

#### Spoken modality

Auditory L1 word recognition has been found to be affected by word frequency, neighbourhood density (usually operationalised as the number of existing words that can be formed by changing one phoneme in the stimulus) and word length. Postulating the null hypothesis that *Lx* word processing draws on the same resources as L1 word processing, Van Heuven (2008) speculates that these factors may also be relevant to auditory *Lx* cognate comprehension.

First, high-frequency L1 words are generally recognised faster than low-frequency words (for references, see Luce and Pisoni, 1998, p. 2 and p. 4; Norris, 2006, p. 327; and Norris and McQueen, 2008, p. 369). Various auditory word recognition models accommodate this finding by modelling word frequency as a proxy of the lexical candidates' prior probability of occurrence or state of activation (e.g. LOGOGEN, Morton, 1969; the Neighborhood Activation Model, Luce and Pisoni, 1998; and Shortlist B, Norris and McQueen, 2008). Similarly, the frequencies of an *Lx* stimulus's cognates in known languages may serve as proxies of the listener's prior expectations when decoding an *Lx* stimulus. Stimuli with highly frequent cognates may therefore get recognised more accurately. Moreover, Kellerman (1983) argued that language users are less apt to transfer known but "psycholinguistically marked" (p. 117) structures to a target language and suggested that these include low-frequency structures. On the account that cognate guessing requires language transfer, then, frequency effects in cognate guessing may come about due to the readers' or listeners' reluctance to consider the possibility that low-frequency items can serve as useful transfer bases.

Second, L1 words with many neighbours are more difficult to recognise than words with few neighbours due to the higher degree of lexical competition caused by perceptually similar words. The frequency of these neighbours, too, is a predictor of the ease of auditory L1 word recognition: words with high-frequency neighbours are more difficult to recognise than words with low-frequency neighbours as high-frequency neighbours are stronger lexical competitors than low-frequency neighbours (Luce and Pisoni, 1998). On the basis of such L1 findings, Van Heuven (2008) hypothesises cross-linguistic neighbourhood density, i.e. the number of

L1 (or, more generally, L1, L2, . . . , L $n$ ) words that can be formed by changing one segment in the L $x$  stimulus, to be negatively correlated with the ease with which an L $x$  stimulus can be decoded. I add that, by the same token, neighbourhood frequency may similarly be a relevant factor in L $x$  stimulus comprehension in that L $x$  stimuli with mainly low-frequency L1, L2, . . . , L $n$  neighbours might be more readily understood than those with mainly high-frequency neighbours.

Third and last, word length, which is related to neighbourhood density and word frequency in that shorter words tend to have more neighbours and frequent words tend to be fairly short, is a co-determiner of L1 word recognition: by and large, longer words contain more phonologically redundant information than do shorter words and are therefore less affected by distortion (for references, see Kürschner et al., 2008, p. 88). Longer L $x$  stimuli, too, contain more information on the basis of which their known cognates can be retrieved, hypothetically increasing their intelligibility.

The importance of these three factors in spoken L $x$  stimulus comprehension was investigated empirically by Kürschner et al. (2008). They found that word length was one of the better—but nonetheless modest—predictors of stimulus comprehension (increase in pseudo- $R^2$ : 0.04–0.06), whereas neighbourhood density and corpus frequency contributed only minimally to their model. In my analyses, I will consequently include word length as a potential predictor. I will also include the corpus frequencies of the target stimuli's German, English and French cognates as potential predictors in order to verify whether Kürschner et al.'s (2008) null result can be replicated. Cross-linguistic neighbourhood density, by contrast, will not be considered as a potential predictor. If cross-linguistic neighbours competed with the correct translation-equivalent cognates in L $x$  cognate guessing tasks, one would expect to find many responses featuring L1, L2, . . . , L $n$  neighbours of the L $x$  stimuli. In the present dataset, common incorrect responses almost never differed in merely one segment from the stimuli.<sup>24</sup> The operationalisation of neighbours as words differing in only one segment from the stimulus may therefore be too rigid to be of use when modelling cognate guessing responses, whereas more lenient operationalisations are unlikely to produce sensible estimates.

---

<sup>24</sup>The only exceptions are the answers *eng* 'narrow' and *ich* (Bernese German: [ɪg]) 'I' for [ɛg] (*ägg* 'egg') with 46 and 15 responses, respectively.

## Written modality

To my knowledge, the effects of word frequency, neighbourhood density and word length on *Lx* stimulus comprehension have never been investigated in the written modality. By analogy with Van Heuven's (2008) hypothesis that these factors are at play in the spoken modality, however, one may submit that they are also relevant in the written modality for the same reasons. First, word frequency is a predictor of the ease with which written L1 words are recognised (for references, see Norris, 2006, p. 327). Second, written L1 words with many orthographic neighbours (operationalised as the number of existing words that can be formed by changing one letter in the stimulus) are harder to recognise than words with few orthographic neighbours (for references, see Norris, 2006, p. 341). Third, longer *Lx* words arguably contain more redundant information than do shorter words in the written modality as well, by virtue of which they may be easier to understand. Of these variables, I will investigate the roles of word frequency and word length on isolated written *Lx* word comprehension for reasons identical to those outlined above.

### 9.1.7 Cross-modality influences

In literate participants, L1 auditory lexical processing is affected by orthographic knowledge (e.g. Jakimik et al., 1985; Peereman et al., 2009; Perre and Ziegler, 2008; Ziegler et al., 2004). If *Lx* processing draws on the same resources as L1 processing, as per Van Heuven's (2008) null hypothesis, it is conceivable that auditory *Lx* decoding is likewise influenced by orthographic knowledge. Schüppert et al. (2010, see Schüppert, 2011, Ch. 6) provide neurological evidence supporting this possibility. Behavioural evidence that listeners make use of their L1 orthographic knowledge when decoding auditory *Lx* stimuli stems from a correlational study by Doetjes and Gooskens (2009) in which 54 Danish pupils translated 86 isolated Swedish cognates. Doetjes and Gooskens (2009) computed both ordinary phonetic Levenshtein distances (such as described above) and Levenshtein distances adjusted for a possible advantageous use of L1 orthography.<sup>25</sup> The adjusted Levenshtein distances

<sup>25</sup>The difference can be illustrated by means of the Danish–Swedish cognate pair [hɔnʔ] *hånd*–[hand] *hand* 'hand'. Although the alignment [hɔnʔ]–[hand] shows an overlap in only two segments, Danes are familiar with the grapheme–phoneme

correlated slightly more strongly with comprehension scores ( $r = -0.63$ ) than did Levenshtein distances based on phonetic transcriptions alone ( $r = -0.54$ ). This higher correlation coefficient may indicate that Danes make use of their L1 phone-to-grapheme mappings when listening to Swedish, but the effect seems to be small. Kürschner et al. (2008) likewise found that a possible advantageous use of L1 orthography helped to account only minimally for differences in comprehension scores of 347 Swedish cognates by Danish pupils. For now, I will leave the possible effect of orthographic compensation strategies in Lx comprehension out of consideration in the present quantitative analyses.

As regards visual cognate guessing tasks, Möller and Zeevaert (2010) observed that participants tend to pronounce the Lx stimuli that they are presented to themselves (either out loud or merely by going through the articulatory motions) on the basis of familiar or assumed grapheme-phoneme correspondences and let these self-pronunciations guide their guesses. I can corroborate this finding from anecdotal observations in the present study. It should be pointed out that these self-articulations do obviously not necessarily conform to the stimuli's actual pronunciations. This allows for substantial inter- and intra-individual variability in the self-articulations: not all participants will necessarily articulate a given written stimulus in the same way and one and the same participant may even consider multiple plausible articulations. Consequently, the effect of such self-pronunciations seems to be too complex to model adequately in quantitative analyses for now and I will not consider it here.

### 9.1.8 Summary and implications

To summarise, readers and listeners have been proposed to be sensitive to various item-related properties when taking part in receptive multilingualism or trying to guess the meaning of cognates in a related language. Unfortunately, not all of these factors can be considered as variables in this study due to small cell sizes ('exotic' properties, lexical stress differences) or due to difficulties linked to their operationalisation (neighbourhood density, cross-modality influences). Still, five kinds of

---

correspondence  $d$ –[d] in their L1. They may thus be able to infer the meaning of Swedish [hand] more successfully than what one would expect on the basis of the phonetic discrepancies by going via their L1 orthography. In the Levenshtein computations adjusted for orthography use, [hɔnʔ]–[hand] therefore overlapped in three segments rather than in only two.

factors remain whose impact on written and spoken cognate guessing can be assessed in this study: (a) overall formal distance between the stimulus and its L1, L2, . . . , L $n$  cognates, (b) the importance of consonants, (c) the importance of word beginnings, (d) cognate frequency and (e) stimulus length.

Of these variables, overall formal distance between the stimulus and its L1, L2, . . . , L $n$  cognates is by far the most extensively researched one and its effect on cognate guessing accuracy can be expected to be negative. As I discussed above, however, several studies indicated that participants in receptive multilingualism and cognate guessing tasks are not equally sensitive to all word parts. Specifically, they may be relatively more sensitive to consonants and to word beginnings. This would require models of cognate guessing to weight consonants and word beginnings more heavily than other word parts. In what follows, I will provide a quantitative assessment of whether weighting consonants and word beginnings more heavily is indeed necessary to account for between-item differences in terms of cognate guessing success. My reasoning is that, if the participants are indeed more sensitive to consonants and word beginnings, then measures of consonantal and word-initial distance should offer explanatory power over and beyond a measure of overall formal distance in which consonants and word-initial discrepancies between the stimuli and their L1, L2, . . . , L $n$  cognates are not weighted more heavily. Finally, the effects of cognate frequency and stimulus length on cognate guessing success have thus far only been investigated quantitatively by Kürschner et al. (2008) and only in the spoken modality. In what follows, I will therefore provide an additional quantitative assessment of their importance in both modalities.

Although the following analyses should shed more light on the item-related factors to which cognate guessers are sensitive, the eventual goal is to investigate whether and how their sensitivity to the relevant factors changes with age. Before I can address this question, however, the item-related variables that are truly predictive of cognate guessing success in this sample of participants should first be identified. This will be the objective of the remainder of this chapter.

A final but highly important conclusion of the literature survey above is that cognate guessers seem to be sensitive not only to item-related properties pertaining to their L1, even if it is closely related to the L $x$ , but also those pertaining to their L2, . . . , L $n$ , as demonstrated by Berthele

(2011). This ties in with the findings regarding inter-individual differences in cognate guessing skills discussed in Section 2.1.2 on page 19 as well as with the findings of the present study discussed in Part II. Nevertheless, foreign language knowledge is not typically taken into account when modelling between-item differences in cognate guessing success. In what follows, I will attempt to overcome this incongruity by also considering item-related properties with respect to the participants' most common L2s and L3s, i.e. French and English. Unfortunately, however, one of the participants' L1s, their Swiss German dialect, cannot be considered to this end given the absence of an agreed-upon orthography and word frequency lists. It is therefore not clear how item-related properties pertaining to Swiss German could be operationalised.

## 9.2 Quantification of predictors

### 9.2.1 Formal distance

#### Overall phonetic distances

The overall phonetic distances between the spoken Swedish stimuli and their cognates were computed by means of the Levenshtein algorithm described in Section 9.1.1. In order to compute these Levenshtein distances, I first transcribed the stimuli and cognates (as presented in Table A.2 on page 192) phonetically. The Swedish stimuli were transcribed perceptually using standard Swedish phonological symbols as used in *Svenska språknämndens uttalsordbok* (Garlén, 2003) and by Engstrand (1990), see Table A.2. For the German, English and French cognates, I used the transcriptions in the Duden *Aussprachewörterbuch* (Mangold, 2005), the *Longman Pronunciation Dictionary* (Wells, 2000) and the *Nouveau Petit Robert* (Rey-Debove and Rey, 1993), respectively. Stress and tonal word accents were not transcribed.

These dictionary transcriptions suffer from a major problem, however: they are based on language-specific traditions and do not lend themselves well to cross-linguistic comparison. Concretely, some phones showing considerable phonetic overlap are transcribed using different symbols, e.g. German [ɛ] and English [e] (see Schmitt, 2007). To improve the interlingual comparability of the transcriptions, I rewrote them using an ad-hoc phone inventory before computing the Levenshtein distances.

The main conversion rules are presented in Table 9.1 on the next page for vowel phones (see also Figure 9.3 on page 126) and Table 9.2 on page 125 for consonant phones. Though necessarily subjective to a certain degree, these conversion rules allow the grouping together of phones that are similar to each other acoustically (e.g. the vowels subsumed into the  $\varepsilon$  category) or phonotactically (e.g. the phones subsumed into the  $r$  category). In addition, I adhered to the following principles when recasting the transcriptions:

- Non-syllabic vowels, e.g. [j] in German [kɔrup'tʃi:ɔn] (*Korruption* ‘corruption’), were rewritten in their consonantal form. Thus, [j] and [w] were substituted for [j̥] and [w̥], respectively.
- Syllabic consonants, e.g. [n̩] in German ['blaibn̩] (*bleiben*, ‘to stay’) or [l̩] in English ['bɒtl̩] *bottle*, were rewritten as the combination of [ə] followed by the non-syllabic allophone of the consonant in question.
- French nasalised vowels, e.g. [ɛ̃] in [ɛ̃ʒenjœr] (*ingénieur* ‘engineer’), were rewritten as the combination of the non-nasalised variant of the vowel in question followed by [n].
- The German sound [ɐ], the typical standard German realisation of /ər/ in words such as ['fɛnstɐ] (*Fenster* ‘window’), was rewritten as [ər], a common Swiss realisation (Christen et al., 2010, pp. 155–158).
- English transcriptions were primarily based on the RP pronunciations provided by Wells (2000), but full rhoticity was assumed. Rhotic vowels, e.g. [ɜː], were subsequently rewritten as the combination of their non-rhotic variants followed by [r]. Thus, *burn* was transcribed not as [bɜːn] but as [bɜːrn] ([bɛːrn] after applying the relevant conversion rule). Moreover, the (General American) notation [oʊ] rather than the (RP) notation [əʊ] was used to represent the GOAT diphthong.

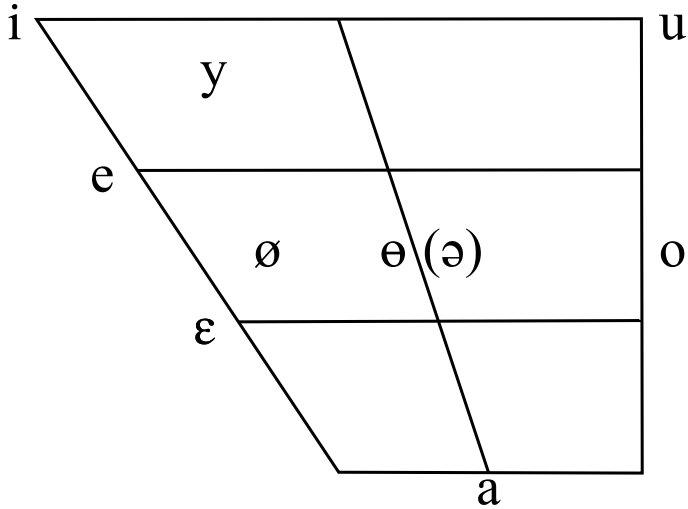
Following these principles, all stimuli and their cognates could satisfactorily be transcribed, though further modifications may be necessary if this ad-hoc phone inventory is to be applied to other stimulus sets. In a last step, I removed length markers and split up diphthongs and affricates into their constituent parts.

**Table 9.1:** Ad-hoc vowel phones used for the Levenshtein distance computations and the Swedish, German, English and French phones subsumed into them.

Ad-hoc phone	Description	Swedish	German	English	French
i	close, front, unrounded	i, ɪ	i, ɪ	i, ɪ	i
y	close, front, rounded	y, ʏ, ʎ	y, ʏ	—	y
e	close mid, front, unrounded	e	e	—	e
ø	mid, front, rounded	ø, œ	ø	—	ø
ɛ	open mid, front, unrounded	ɛ, æ	ɛ	e	ɛ
a	open, unrounded	a, ɑ	a	æ, ɑ	a
o	mid, back, rounded	o, ɔ	o, ɔ	ɔ, ɒ	o, ɔ
u	close, back, rounded	u, ʊ	u, ʊ	u, ʊ	u
ə	mid, central, stressed	ə	—	ʌ, ɜ	—
ə	mid, central, unstressed	ɛ, œ (unstressed)	ə	ə	ə

**Table 9.2:** Ad-hoc consonant phones used for the Levenshtein distance computations and the Swedish, German, English and French phones subsumed into them.

Ad-hoc phone	Description	Swedish	German	English	French
p / b	labial plosives	p / b	p / b	p / b	p / b
t / d	coronal plosives	t, t̥ / d, d̥	t / d	t / d	t / d
k / g	dorsal plosives	k / g	k / g	k / g	k / g
m	labial nasal	m	m	m	m
n	coronal nasal	n, ŋ	n	n	n
ŋ	dorsal nasal	ŋ	ŋ	ŋ	ŋ
f / v	labial fricatives	f / v	f / v	f / v	f / v
θ / ð	dental fricatives	— / —	— / —	θ / ð	— / —
s / z	alveolar fricatives	s / —	s / z	s / z	s / z
ʃ / ʒ	postalveolar fricatives	ʃ, ʒ / —	ʃ / ʒ	ʃ / ʒ	ʃ / ʒ
ç	dorsal fricative	ç	ç, x	—	—
h	laryngeal fricative	h	h	h	—
r	r sound	r	r, R	r	R
l	lateral approximant	l, ʎ	l	l	l
j	palatal approximant	j	j	j	j
w	velar approximant	—	—	w	w, ʋ



**Figure 9.3:** Ad-hoc vowel inventory used for the phonetic Levenshtein computations.

The Levenshtein distances between the transcriptions of the stimuli and their cognates were computed using an algorithm that weighted all operations (insertions, deletions, substitutions) equally. Thus, substituting [i] for [e] cost the same as substituting [i] for [a]. Only vowel–vowel and consonant–consonant mappings were allowed. Following Heeringa’s (2004) suggestion, the raw distances were length-normalised by dividing them by the length of the longest lowest-cost alignment. For stimuli lacking a cognate in a given source language, the respective Levenshtein distance was arbitrarily set to 1. In the three cases in which the stimulus had two cognates in German (*fråga*, *larm*, *öst*), I used the cognate associated with the lower Levenshtein distance for these computations (*Frage*, *Alarm* and *Ost*, respectively). Every target word thus has three associated Levenshtein distances, one for each of the three potential supplier languages under consideration (German, English and French).

## Overall orthographic distances

The overall orthographic distances between the written Swedish stimuli and their cognates were likewise computed by means of the Levenshtein algorithm. Before the Levenshtein distance computations, the orthographic strings representing the visual stimuli and their translation-equivalent cognates were converted to lowercase. The German letter  $\beta$ , which is not common in written standard German in Switzerland, was represented as *ss*.

The Levenshtein distances between the orthographic strings were computed using an algorithm that weighted all operations (insertions, deletions, substitutions) equally. Diacritical differences, e.g. between  $\ddot{o}$  and *o* in *öppna*–*open*, did not count towards the Levenshtein distance.<sup>26</sup> Only vowel–vowel and consonant–consonant mappings were allowed. For this purpose, the graphemes *a*, *e*, *i*, *o*, *u* and *y* as well as their versions with diacritics were considered vowels; all other graphemes were considered consonants. The raw distances were length-normalised by dividing them by the length of the longest lowest-cost alignment. For stimuli lacking a cognate in a given source language, the respective Levenshtein distance was arbitrarily set to 1.

## Consonantal distances

I operationalised the consonantal distances between the Swedish stimuli and their German, English and French cognates similarly to Gooskens et al. (2008). I first counted the number of consonant operations in the full Levenshtein alignments described above. These counts were then length-normalised by dividing them by the length of the full Levenshtein alignment. Consider, for instance, Figure 9.2(b) on page 108. In order to transform [sykəl] into [tsyklʊs], two consonantal as well as two vocalic operations are needed (normalised Levenshtein distance: 0.50). Length-normalising the consonantal count yields a consonantal distance of

---

<sup>26</sup>Levenshtein distances for which aligned graphemes differing solely in their diacritics received half the weight of an ordinary operation were for all practical purposes identical to the Levenshtein distances for which diacritics were ignored. For the three language pairs (Swedish–German, Swedish–English and Swedish–French), the correlation coefficients between both Levenshtein measures ranged between 0.991 and 0.997.

$2 \div 8 = 0.25$ .<sup>27</sup> The consonantal Levenshtein distance for stimuli lacking a cognate in a given source language was arbitrarily set to 1.

### Word-initial distances

In order to assess whether word-initial similarity between  $L_x$  stimuli and their  $L_1, L_2, \dots, L_n$  is particularly important in cognate guessing tasks, I computed the Levenshtein distances between the word beginnings of the stimuli and their German, French and English cognates. Word beginnings were operationally defined as up to and including the first consonant or consonant cluster. Thus, the word beginnings in the cognate pair *avskaffa*–*abschaffen* ‘to abolish’ are *av* and *ab*. Similarly to the consonantal distances, the word-initial distances were length-normalised by dividing them by the length of the overall Levenshtein alignment. In the case of *avskaffa*–*abschaffen*, the word-initial Levenshtein distance thus equals  $1 \div 10 = 0.10$  (there are ten slots in the overall alignment).<sup>28</sup> The word-initial Levenshtein distance for stimuli lacking a cognate in a given source language was arbitrarily set to 1.

---

<sup>27</sup>An alternative approach is to length-normalise these counts by the number of consonant alignments in the full Levenshtein alignment. For the [svkɛl]–[tsyklʊs] example, this yields a consonantal distance of  $2 \div 5 = 0.40$  (there are five slots associated with consonants). These alternative consonantal distances are very strongly correlated with the ones described in the main text, however: broken down by modality (written–spoken) and language pair (Swedish–German, Swedish–English and Swedish–French), the six correlations range from 0.91 to 0.99 (median: 0.98). The results regarding consonantal Levenshtein distances presented in this chapter are thus robust with respect to the normalisation procedure chosen, and I will not discuss these alternative consonantal Levenshtein distances any further lest they clutter up the main text.

<sup>28</sup>An alternative approach is to length-normalise the word-initial Levenshtein distances by dividing them by the length of the word-initial alignment only (compare Note 27). According to this approach, the word-initial Levenshtein distance between *avskaffa* and *abschaffen* equals  $1 \div 2 = 0.50$  (there are two slots in the word-initial alignment). These alternative word-initial Levenshtein distances are strongly correlated with the ones described in the main text: broken down by modality (written–spoken) and language pair (Swedish–German, Swedish–English and Swedish–French), the six correlations range from 0.64 to 0.95 (median: 0.91). The results discussed in this chapter are consequently robust with respect to the normalisation procedure chosen, and I will not further discuss these alternative word-initial Levenshtein distances.

### 9.2.2 Cognate frequencies

As cognate frequency measures, I extracted the word frequencies per 1,000,000 words of the German and English translation-equivalent cognates of the Swedish stimuli from the SUBTLEX-DE (Brysbaert et al., 2011) and SUBTLEX-US (Brysbaert and New, 2009) databases, respectively. The SUBTLEX databases can be downloaded free of charge from <http://crr.ugent.be/programs-data/subtitle-frequencies>. Frequencies per 1,000,000 words for the French cognates were extracted from the highly comparable Lexique 3 database (New et al., 2007), which is freely available from <http://www.lexique.org/>. Unlike the SUBTLEX databases, Lexique 3 distinguishes homographs by part-of-speech. For full comparability with the SUBTLEX frequencies, I aggregated the frequencies for homographs. For target words with more than one translation-equivalent cognate, I summed over the frequencies of all translation-equivalent cognates. The German cognate frequency for *fråga* ‘question; to ask’ is therefore the sum of the SUBTLEX-DE frequencies for *Frage* ‘question’ and *fragen* ‘to ask’. All frequency measures were logarithmically transformed in order to prevent items with extremely high frequency counts from exerting undue influence on the analyses. I added 1 to every frequency count in order to deal with zero frequencies, the logarithm of which would otherwise be undefined:  $\log\text{-frequency} \equiv \ln(\text{frequency} + 1)$ .

### 9.2.3 Stimulus length

Stimulus length was quantified as the number of graphemes (for written target words) or phones (for spoken target words) in the Swedish stimuli.

## 9.3 Variable selection

On the basis of previous findings, five classes of predictors were identified that could help to explain between-item differences in cognate guessing accuracy and that have now been quantitatively operationalised: (a) overall Levenshtein distance, (b) consonantal Levenshtein distance, (c) word-initial Levenshtein distance, (d) cognate frequency and (e) stimulus length. In classes (a) to (d), there are three predictors each, one for each potential supplier language under consideration (German, English and

French). Thirteen predictors is obviously too many to consider jointly in a regression that models the comprehension of merely 45 written items or 42 spoken items, a problem that is compounded by the fact that these predictors show moderate to severe collinearity ( $\kappa = 31$  for the written items and 24 for the spoken items; see Baayen, 2008, p. 182). Thus, the first priority is to select from this set of potential predictors the ones that can best and independently explain the between-item differences in cognate guessing accuracy in the data set. In this variable selection process, these between-item differences are represented by the random intercepts from the models presented in Tables 7.1 and 7.2 on pages 87 and 92. I prefer this approach over computing the percentage of correct translations per item mainly for conceptual reasons (For these data, the differences between both approaches are unlikely to be large.): first, unlike such percentages, they are not bound between two values (0 and 100 in the case of percentages), and second, the random intercepts are expressed in log-odds, which reflects the binary nature of the actual outcome variable.

### 9.3.1 Bivariate relationships

One variable selection strategy is to compute the correlation coefficients of the bivariate relationships between the predictors and the by-item random intercepts. Table 9.3 on the next page provides these coefficients, which indicate that the comprehension of the written items correlates most strongly with the Levenshtein distance with respect to German ( $r = -0.43$ ), followed by German cognate frequency ( $r = 0.37$ ) and English cognate frequency ( $r = 0.35$ ). The comprehension of the spoken items correlates most strongly with the Levenshtein distance with respect to German ( $r = -0.66$ ), followed by the consonantal Levenshtein distance with respect to German ( $r = -0.53$ ).

This strategy suffers from three drawbacks, however. First, the collinearity between the predictors is not taken into consideration. The correlation coefficients are therefore likely to overestimate the actual importance of several predictors. Second, the correlation coefficients reflect the strength of the *linear* relationship between the predictors and the outcome variable. Visual data explorations (not reported here), however, suggest the presence of some non-linear bivariate relationships in the data. Third, the item-related predictors might enter into interactions with each other, which are not be detectable in this approach. To

**Table 9.3:** Pearson correlation coefficients of the bivariate relationships between the item-related predictor variables and the by-item random intercepts for written ( $n = 45$ ) and spoken items ( $n = 42$ ).

		Written items	Spoken items
Stimulus length		-0.20	0.04
Full Levenshtein distance	German	-0.43	-0.66
	English	-0.26	-0.26
	French	0.21	-0.27
Consonantal Levenshtein distance	German	-0.16	-0.53
	English	-0.12	-0.19
	French	0.24	-0.20
Word-initial Levenshtein distance	German	-0.11	-0.42
	English	-0.12	-0.15
	French	0.24	-0.21
Cognate frequency (log)	German	0.37	0.04
	English	0.35	0.05
	French	-0.12	0.21

circumvent these drawbacks, I turn to an approach that does not suffer so strongly from them: *conditional variable importance* estimation on the basis of *random forests*.

### 9.3.2 Random forest-based conditional permutation importance

#### Rationale of random forests

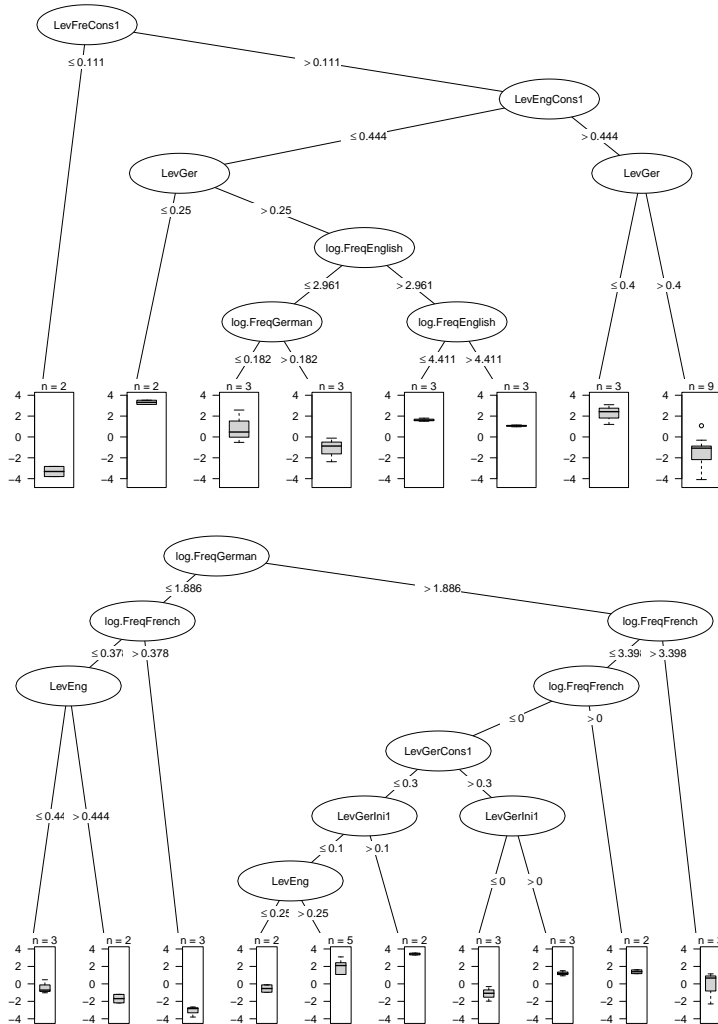
The logic behind random forests and random forest-based variable importance estimation is explained accessibly by Strobl et al. (2008, 2009b), and an introduction geared towards language researchers is provided by Tagliamonte and Baayen (2012); here I will briefly take up the main points. Random forests are ensembles of *classification* or *regression trees*. Classification and regression trees seek to explain the variance in an outcome variable by recursively partitioning the data by means of binary splits so as to reach ever purer (i.e. more uniform with respect to the outcome variable) nodes. If the outcome variable is a nominal variable, the algorithm is called a classification tree; if the outcome variable is continuous, as in the present case, it is called a regression tree.

Regression trees are flexible quantitative tools in that they can easily cope with interacting predictors, non-linearities and a multitude of predictors relative to the number of observations ('large  $p$ , small  $n$  cases'). They do, however, suffer from two severe drawbacks in particular. First, they are highly unstable: small changes in the data can and often do result in dramatically different tree models, which in turn often give rise to radically different interpretations. Second, they are piecewise continuous: even continuous relationships between the outcome variable and the predictor variables are broken down into dichotomies, which may suggest the presence of threshold effects when in fact none exist.

A popular solution to these problems is to grow not one but several hundreds of trees. By randomly resampling from the original set of cases (either with or without replacement), 'new' data sets are created on which new, different trees can be grown. Due to the random fluctuations in the training data, the ensemble as a whole is much more robust than a single tree and the hard-cut boundaries that are characteristic of single trees are smoothed. When all potential predictors are considered at each stage to determine whether and how the nodes can further be purified, this procedure is called *bagging* (Breiman, 1996). In order to grow even more diverse trees, the set of possible predictors can be reduced randomly. For instance, we can specify that at each stage, only four out of 13 predictors are to be taken into consideration. This approach, called *random forests* (Breiman, 2001), "allows predictor variables that were otherwise outplayed by their competitors to enter the ensemble" (Strobl et al., 2008) and may thereby reveal subtle interaction effects that would have remained hidden otherwise.

In Figure 9.4 on the facing page, I provide an example of a small random forest consisting of merely two trees (i.e. `ntree` = 2) grown on two (different) subsamples of 28 cases each sampled randomly from the 45 written items. At each stage, four randomly selected predictors were considered in order to determine the next split (i.e. `mtry` = 4). Note how the trees have radically different shapes.

In the example above, each tree was grown on the basis of 28 out of 45 observations, i.e. 17 observations were not 'seen' by a given tree and did not affect the tree growing process. Such observations are called 'out-of-bag' (OOB) observations. The prediction accuracy of a random forest is estimated by letting each tree 'decide' on the probable outcome



**Figure 9.4:** A small random forest of two trees grown on the data for the written items. Each tree models 28 randomly chosen items and for each split, four randomly selected predictor variables out of 13 were considered (i.e. `mtry = 4`). Note how the trees have radically different shapes.

value of these OOB observations and checking how well the average predicted outcomes match the actual outcomes.

### Variable importance estimation

Larger random forests often produce excellent OOB prediction rates, but unlike single trees, they are difficult to visualise and interpret and are effectively ‘black boxes’. To gain some insight into which variables are important, we can compute several variable importance measures, e.g. the so-called *permutation importance*. This importance measure is derived by computing how much the overall OOB prediction accuracy of the random forest decreases when the values of a given predictor are randomly permuted, thereby breaking the association between the predictor and the outcome variable. The more important a predictor is in a random forest ensemble, the more this permuting will affect its overall prediction accuracy. Such an approach, however, may unduly favour correlated variables, which is why Strobl et al. (2008) proposed the *conditional permutation importance*, for which the intercorrelations between predictor variables are reduced by means of a permutation-based conditioning scheme (see also Strobl et al., 2009a). The variable importance measures thus computed reflect more closely the partial effects of each variable. That said, very strong intercorrelations between predictors may still be difficult to permute away entirely, so that even conditional permutation importances overestimate the effect of highly correlated variables somewhat.

### Settings

To return to the problem at hand, I grew two separate random forests on the data for the written items and for the spoken items using the `cforest()` function in the `party` package (version 1.0-6; Hothorn et al., 2013) for R. The individual trees were grown on subsamples of the data rather than bootstrap samples (i.e. resampling without rather than with replacement) and each subsample consisted of 63.2% of the cases of the original data set (see Strobl et al., 2007). Each forest consisted of 1,000 trees (i.e. `ntree` = 1000) and four randomly selected predictors out of 13 were considered at each split (i.e. `mtry` = 4). The by-item random intercepts were the dependent variables. I then computed the conditional permutation importance measures for the predictors using

party's `varimp()` function. If the significance (i.e. the  $p$ -value) of the association between a given predictor and another covariate was lower than 0.80 (the default), the relevant covariate was included in the predictor's conditioning scheme.

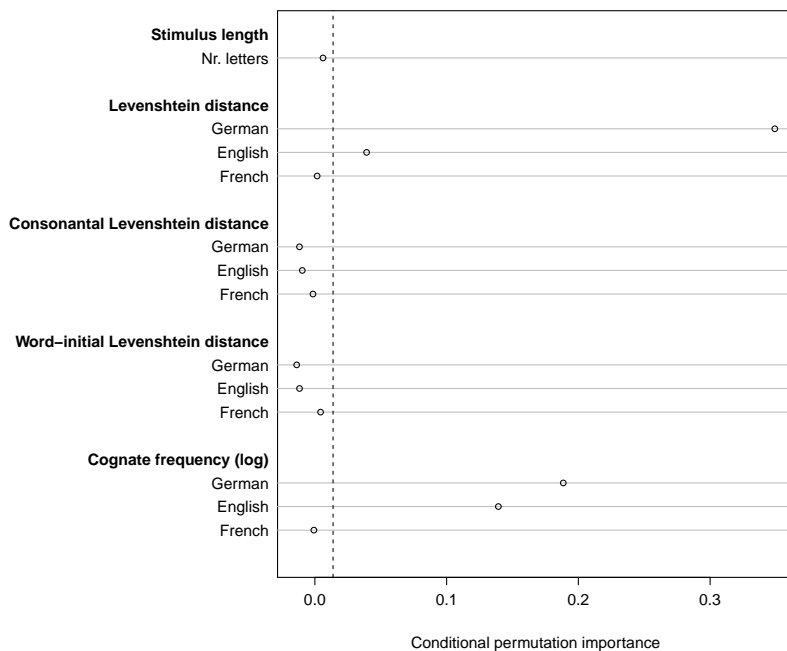
### 9.3.3 Results and discussion

The predictors' conditional permutation importance measures are given in Figure 9.5 on the next page for the written items and in Figure 9.6 on page 137 for the spoken items.<sup>29</sup> A few variables have negative permutation importances, which would indicate that the prediction accuracy of the random forests actually increases when the information contained in these variables is lost due to the permutation algorithm. Prediction accuracy obviously cannot really improve when relevant information is lost, and these negative values can thus only be due to random variation in the permutation algorithm: by sheer luck, random predictor values predict the outcome variable better than the actual values. These negative permutation importances thus provide a conservative indication of how much the permutation importances can deviate from zero due to randomness alone. The dotted vertical lines in Figures 9.5 and 9.6 mark the amplitude of the largest negative permutation scores; variables with permutation scores to the left of these lines can be considered to be effectively irrelevant in the random forests (see Strobl et al., 2009b, p. 339). Note, parenthetically, that absolute variable importances should not be compared with each other across datasets or studies (Strobl et al., 2008): the fact that the importance scores in Figure 9.5 range approximately from 0 to 0.35 and those in Figure 9.6 from 0 to 1.4 does not necessarily merit further discussion.

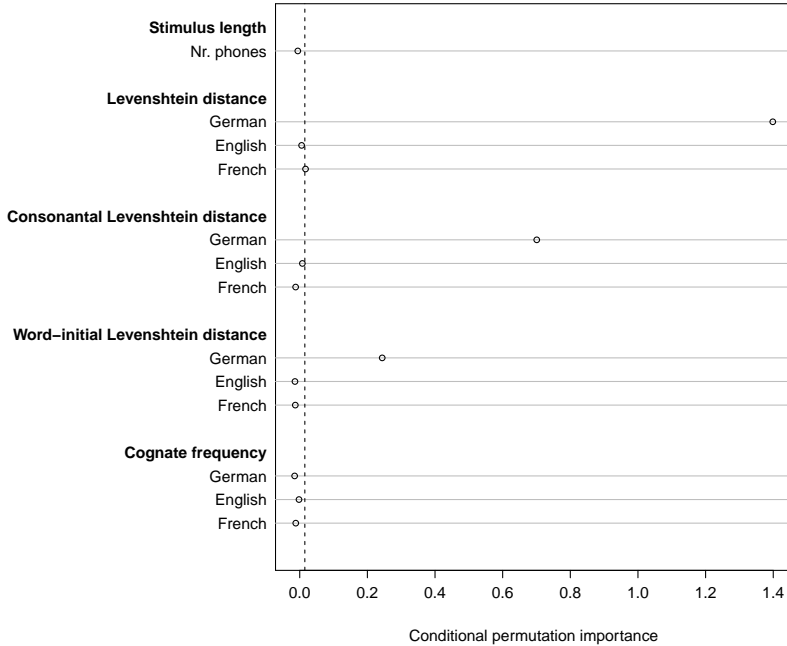
I will briefly discuss the permutation scores in Figures 9.5 and 9.6 with a view towards selecting a handful of variables for the full-fledged analyses; a more in-depth discussion of their implications is postponed to Section 9.5. Clearly, for both written and spoken items, stimulus

---

<sup>29</sup>Note that these variable importance measures are subject to random fluctuations introduced in the forest growing process (random subsamples and random variable selection) and in the permutation process. Consequently, readers who wish to compute these measures for themselves will find that the precise values may differ somewhat from those presented here, particularly the lower values. I ran the algorithms several times, however, and found the relative rankings of the most important variables to be highly similar between runs. Changing the `mtry` and `ntree` parameters did not substantially affect the results either.



**Figure 9.5:** Conditional permutation importances of the predictors in a random forest ( $mtry = 4$ ,  $ntree = 1000$ ) modelling the random intercepts of the written items ( $n = 45$ ). The dotted line indicates the amplitude of the largest negative permutation score (due to randomness). Variables with permutation scores to the left of this line can be considered to be irrelevant in the random forest.



**Figure 9.6:** Conditional permutation importances of the predictors in a random forest (`mtry = 4`, `ntree = 1000`) modelling the random intercepts of the spoken items ( $n = 42$ ). The dotted line indicates the amplitude of the largest negative permutation score (due to randomness). Variables with permutation scores to the left of this line can be considered to be irrelevant in the random forest.

length is an irrelevant factor. Second, all variables with respect to French turn out to have negligible permutation scores in both modalities and can likewise be left out of consideration in the regression analyses. Third, overall Levenshtein distances are much more important than consonantal and word-initial Levenshtein distances in both modalities. The consonantal and word-initial Levenshtein distances do not all hover around 0 in the spoken modality, however. This may indicate either that these variables may explain some variance in cognate translation accuracy over and beyond what can be accounted for by the overall Levenshtein distances or that the correlation between these variables on the one hand and the overall Levenshtein distances on the other hand was too strong to be scrambled away entirely by `varimp()`'s conditioning scheme. Fourth, cognate frequency may play some role in the written modality, but its role in the spoken modality is negligible and does not merit further consideration in the modelling process.

A noteworthy fifth point is that, in the written modality, the variables with respect to English are potentially relevant predictors whenever the corresponding variables with respect to German are identified as potentially relevant predictors, too. This may indicate that the participants tend to draw on their knowledge of both German and English when engaging in a written cognate guessing task. Hypothetically, participants may tend to evaluate the orthography of an  $Lx$  stimulus with respect to both its German and English cognates simultaneously. Put somewhat simplistically, they might tend to prefer whichever of the German and English words considered that shows the smallest orthographic distance to the stimulus as the most plausible translation candidate. This would imply that it is not necessarily the Levenshtein distance with respect to German that is the best predictor of written item comprehension, but rather whichever of the German and English Levenshtein distances that happens to be the lower one. Furthermore, it need not be the language-specific frequencies that play a role in cognate guessing; when a stimulus has cognates in both German and English, it may be the (possibly weighted) sum of both cognates' frequencies that is the better predictor of cognate guessing accuracy.

In order to take these possibilities into account, I computed 'Germanic' Levenshtein and frequency variables. For each stimulus, the overall Germanic Levenshtein distance equals whichever of the German or English Levenshtein distance is the lower one. The consonantal and

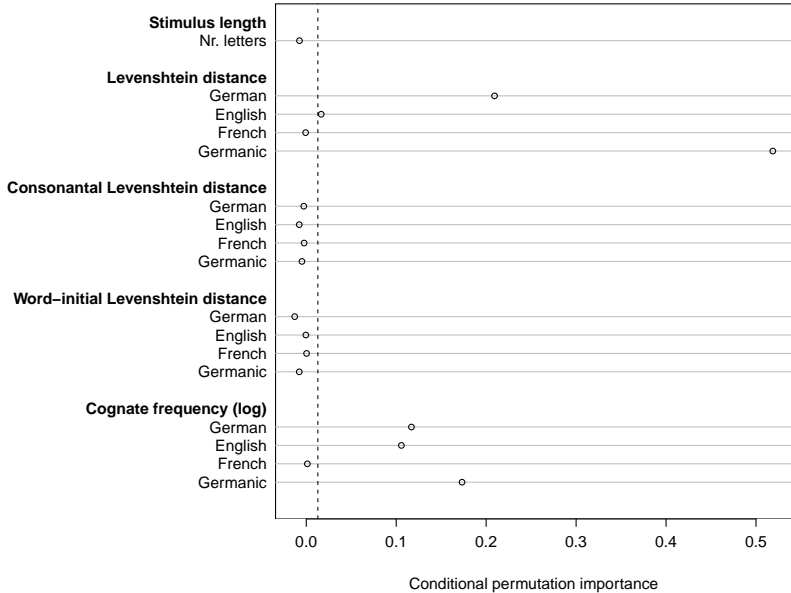
word-initial Germanic Levenshtein distances are derived from the alignments associated with this overall Germanic Levenshtein distance. The Germanic cognate frequency is the mean of the stimulus's German and English cognate frequencies, which was then logarithmically transformed ( $\ln(\text{meanfrequency} + 1)$ ). These four Germanic variables were added to the original 13 in order to grow another random forest (`mtry = 4`, `ntree = 1000`), on the basis of which a new set of conditional permutation importances was computed (see Figure 9.7 on the next page). Clearly, overall Germanic Levenshtein distance and Germanic cognate frequency emerge as more important than their language-specific counterparts. As per the discussion on page 134, the other variables with permutation scores to the right of the vertical line may either be genuine predictors of written item comprehension or be parasitic on the stronger predictors.

In summation, the random forest-based conditional permutation importances indicate that overall Germanic Levenshtein distance is the best predictor of written item comprehension, whereas overall Levenshtein distance with respect to German emerged as the best predictor of spoken item comprehension. Furthermore, Germanic cognate frequency may be a useful predictor of written item comprehension as well. In order to render the effects of these variables more interpretable, I fitted them in regression models, which I will present in the next section.

## 9.4 Regression modelling

### 9.4.1 Written items

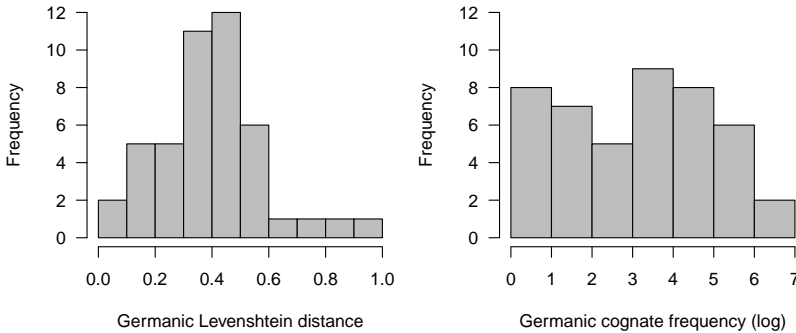
The effects of overall Germanic Levenshtein distance and Germanic cognate frequency on cognate guessing accuracy were first fitted in a generalised additive model with crossed random intercepts for stimuli and participants. This revealed that both effects were approximately linear. Adding the other variables that have non-zero conditional permutation importances in Figure 9.7 (i.e. German and English Levenshtein distance as well as German and English cognate frequency) did not improve the model fit. This suggests that the correlations between these variables and their Germanic counterparts were too strong to be scrambled away entirely by the conditional permutation algorithm. In conclusion, cognate guessing accuracy in the written modality seems to be affected



**Figure 9.7:** Conditional permutation importance of the predictors in a random forest ( $mtry = 4$ ,  $ntree = 1000$ ) modelling the random intercepts of the written items ( $n = 45$ ). The variables pertaining to German and English were collapsed into ‘Germanic’ predictors. The dotted line indicates the amplitude of the largest negative permutation score (due to randomness). Variables with permutation scores to the left of this line can be considered to be irrelevant in the random forest.

**Table 9.4:** Descriptive statistics of the item-related variables affecting written cognate guessing accuracy computed with respect to the 45 written target words.

	Range		Median	Mean	SD
	Lower	Upper			
Germanic Levenshtein distance	0.00	1.00	0.40	0.42	0.19
Germanic cognate frequency (log)	0.00	6.14	3.20	3.09	1.80



**Figure 9.8:** Histograms of the item-related variables affecting cognate guessing accuracy for written items ( $n = 45$ ).

primarily by overall Germanic Levenshtein distance and Germanic cognate frequency. The descriptive statistics of these two key variables are presented in Table 9.4 and their univariate distributions are plotted in Figure 9.8. Since the effects of these variables are approximately linear, they were modelled in a generalised linear mixed-effects model.

I added overall Germanic Levenshtein distance and (log-transformed) Germanic cognate frequency as fixed effects to the model summarised in Table 7.1 on page 87. Both variables were centred at their sample means following Baayen (2008, pp. 254–255). Additionally, model comparisons favoured the inclusion of by-participant random slopes for these effects. The resultant model is summarised in Table 9.5 on the following page. The fixed effects of overall Germanic Levenshtein distance and Germanic cognate frequency are presented visually in Figure 9.9. Table 9.5 and

**Table 9.5:** Generalised (logistic) mixed-effect model modelling translation accuracy on written target words in function of item- and participant-related predictors. (a) Fixed effects, their two-tailed significance and their effect sizes. (b) Modelled standard deviation of the random effects ( $\hat{\sigma}$ ). All covariates were centred at their sample means. Parameters and effect sizes are expressed in log-odds and are reported to two significant digits.

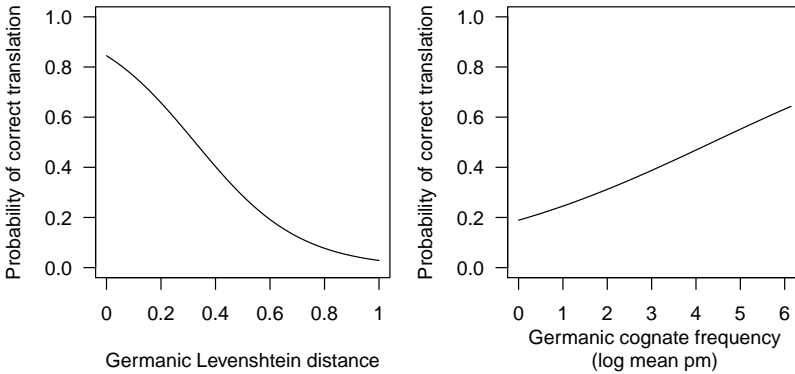
(a) Fixed effects			
	Estimate $\pm$ SE	$p$	Effect size $\pm$ SE <sup>a</sup>
Intercept	-0.92 $\pm$ 0.24	<0.001	
$\geq 1$ correct profile word translations	0.49 $\pm$ 0.17	0.003	0.49 $\pm$ 0.17
Number of foreign languages	0.17 $\pm$ 0.07	0.016	0.69 $\pm$ 0.29
English proficiency	0.20 $\pm$ 0.06	<0.001	1.5 $\pm$ 0.4
WST score	0.078 $\pm$ 0.014	<0.001	2.9 $\pm$ 0.5
Germanic Levenshtein distance	-5.2 $\pm$ 1.2	<0.001	5.2 $\pm$ 1.2
Germanic cognate frequency (log)	0.33 $\pm$ 0.12	0.007	2.0 $\pm$ 0.8

<sup>a</sup> Effect sizes were computed as the largest absolute difference in the outcome variable (in log-odds) when the predictor variable is allowed to vary along its range. See Table 7.1 for an example.

(b) Random effects	
	$\hat{\sigma}$
Random intercept by participant	0.69
Random slope for Germanic Levenshtein distance by participant	1.5
Random slope for Germanic cognate frequency (log) by participant	0.097
Random intercept by item	1.5
Random slope for English proficiency by item	0.20
Random slope for WST score by item	0.057

Figure 9.9 indicate that translating written  $Lx$  cognates becomes more difficult as the formal discrepancies between the  $Lx$  stimuli and their German and English cognates increase and more easy if the  $Lx$  stimuli have high-frequency German and English cognates. The fixed effects of the participant-related variables are essentially identical to those plotted in Figure 7.1 on page 88 and were not plotted again.

One of the advantages of the mixed-effects approach is that the relative impact of participant- and item-related predictors can now straightforwardly be compared in the same model whilst accounting for idiosyncratic differences in a principled way using random slopes. In the



**Figure 9.9:** Partial fixed effects of the GLMM modelling the translation accuracy on written target items in terms of item- and participant-related predictors. Only the effects of the item-related predictors were plotted; for the effects of the participant-related predictors, see Figure 7.1 on page 88. Nominal variables not included in a given plot were fixed at their modes, and continuous variables not included in a given plot were fixed at their medians.

**Table 9.6:** Descriptive statistics of the item-related variable affecting spoken cognate guessing accuracy computed with respect to the 42 written target words.

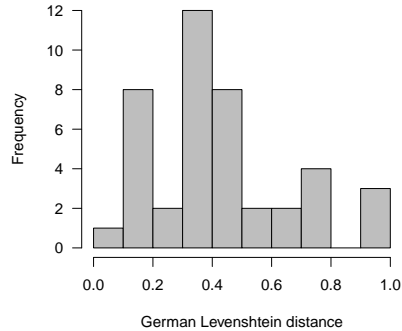
	Range		Median	Mean	SD
	Lower	Upper			
German Levenshtein distance	0.00	1.00	0.40	0.44	0.24

present data set, overall Germanic Levenshtein distance, an item-related variable, is clearly the most important predictor of all with an effect size of  $5.2 \pm 1.2$  log-odds, whereas Germanic cognate frequency is also a respectable predictor (ES:  $2.0 \pm 0.8$ ) in comparison with participant-related variables such as WST score (ES:  $2.9 \pm 0.5$ ) and English proficiency (ES:  $1.5 \pm 0.4$ ).

### 9.4.2 Spoken items

The effect of overall German Levenshtein distance on cognate guessing accuracy in the spoken modality was first modelled in a generalised additive model with crossed random intercepts for stimuli and participants. This model revealed that the effect could be modelled approximately linearly. Other variables that have non-zero conditional permutation importances in Figure 9.6 (i.e. consonantal and word-initial German Levenshtein distance) did not contribute to the model fit, suggesting that their non-zero permutation importances are the result of their collinearity with overall German Levenshtein distance. The summary statistics of the German Levenshtein distance variable are presented in Table 9.6 and its univariate distribution is plotted in Figure 9.10 on the next page.

Since the effect of German Levenshtein distance on cognate guessing accuracy in the spoken modality could be modelled approximately linearly, it was added to the generalised linear mixed-effects model presented in Table 7.2 on page 92. The Levenshtein variable was centred at its sample mean following Baayen (2008, pp. 254–255). Model comparisons did *not* favour the inclusion of a random slope parameter modelling between-participant differences in the effect of the overall German Levenshtein variable. The model is summarised numerically in Table 9.7 on page 146 and the effect of German Levenshtein distance is presented



**Figure 9.10:** Histograms of the item-related variable affecting cognate guessing accuracy for spoken items ( $n = 42$ ).

visually in Figure 9.11 on page 147. The effects of the participant-related variables were essentially identical to those plotted in Figure 7.2 on page 93 and were not plotted again.

Table 9.7 and Figure 9.11 unsurprisingly indicate that cognate guessing accuracy in the spoken modality drops sharply as the phonetic Levenshtein distance between the stimuli and their German cognates increases. In fact, with an effect size of  $6.8 \pm 1.2$  log-odds, the Levenshtein variable is three to four times more important in predicting cognate guessing accuracy than the participant-related variables of Raven score (ES:  $1.9 \pm 0.4$ ), English proficiency (ES:  $1.4 \pm 0.4$ ) and WST score (ES:  $1.3 \pm 0.5$ ).

## 9.5 Discussion

The goal of this chapter was to find the set of predictors that most parsimoniously explains the between-item variance in  $Lx$  cognate guessing accuracy in order to then investigate whether the strength of the effects of these predictors vary systematically in function of the participants' age. The ultimate goal, then, is to gain further insight into how cognate guessing skills change with age. Still, the analyses presented in this chapter also have value of their own as they provide a quantitative evaluation

**Table 9.7:** Generalised (logistic) mixed-effect model modelling translation accuracy on spoken target words in function of item- and participant-related predictors. Panel (a): Fixed effects, their two-tailed significance and their effect sizes. Panel (b): Modelled standard deviation of the random effects ( $\hat{\sigma}$ ). All covariates were centred at their sample means. Parameters and effect sizes are expressed in log-odds and are reported to two significant digits.

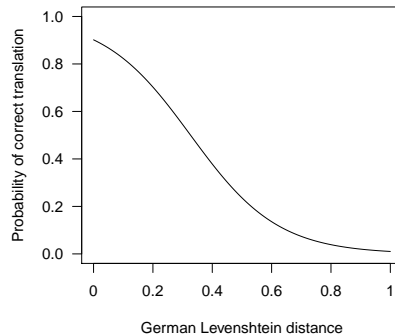
(a) Fixed effects

	Estimate $\pm$ SE	$p$	Effect size $\pm$ SE <sup>a</sup>
Intercept	-1.1 $\pm$ 0.3	<0.001	
English proficiency	0.18 $\pm$ 0.06	0.001	1.4 $\pm$ 0.4
WST score	0.036 $\pm$ 0.015	0.013	1.3 $\pm$ 0.5
Raven score	0.053 $\pm$ 0.011	<0.001	1.9 $\pm$ 0.4
Backward digit span	-0.086 $\pm$ 0.041	0.035	0.86 $\pm$ 0.41
German Levenshtein distance	-6.8 $\pm$ 1.2	<0.001	6.8 $\pm$ 1.2

<sup>a</sup> Effect sizes were computed as the largest absolute difference in the outcome variable (in log-odds) when the predictor variable is allowed to vary along its range. See Table 7.1 for an example.

(b) Random effects

	$\hat{\sigma}$
Random intercept by participant	0.68
Random intercept by item	1.9
Random slope for English proficiency by item	0.13
Random slope for WST score by item	0.054
Random slope for Raven score by item	0.029



**Figure 9.11:** Partial fixed effects of the GLMM modelling the translation accuracy on spoken target items in terms of item- and participant-related predictors. Only the effect of the item-related predictor was plotted; for the effects of the participant-related predictors, see Figure 7.2 on page 93. Continuous variables not included in a given plot were fixed at their medians.

of the role of several item-related variables in cognate guessing. The results therefore merit some brief discussion.

### 9.5.1 Variables considered

#### Overall formal distance

In consonance with several previous findings (Doetjes and Gooskens, 2009; Gooskens et al., 2011; Kürschner et al., 2008; Van Bezooijen and Gooskens, 2005a), the overall phonetic or orthographic distances between the Swedish stimuli and their L1 (German) cognates (as computed by means of the Levenshtein algorithm) were the strongest of the predictors of item comprehension that were initially considered, both in the written and in the spoken modality. This is clear both from the Pearson correlation coefficients presented in Table 9.3 on page 131 and the conditional permutation importances presented in Figures 9.5 and 9.6 on pages 136 and 137. The conditional variable importances in Figure 9.5, however, indicated that the orthographic distances with respect to English affect written item comprehension somewhat as well. I therefore

computed the ‘Germanic’ orthographic Levenshtein distance to each Swedish stimulus. This Germanic Levenshtein distance equals whichever of the German or English Levenshtein distance to each Swedish item is the lower one (similarly to Berthele and Lambelet, 2009) and proved to be an even stronger predictor of written item comprehension as is evident from the conditional variable importances presented in Figure 9.7 on page 140. Orthographic distance with respect to French, by contrast, did not emerge as a relevant predictor of written item comprehension, and no multilingual measure of orthographic distance including French was therefore included into the regression models. The conditional variable importances for the spoken items (Figure 9.6) did not reveal the phonetic Levenshtein distances with respect to English or French to be relevant predictors of item comprehension, and I consequently did not collapse the Levenshtein distances with respect to German, English and French into a single multilingual predictor variable.

As expected, the effects of the Levenshtein variables are negative in both modalities: greater formal discrepancies between the  $Lx$  stimuli and their cognates in a known language are associated with a lesser degree of cognate guessing success. In fact, Levenshtein distance is not only the most important item-related predictor of cognate guessing success; the effect sizes given in Tables 9.5 and 9.7 suggest it to be a stronger predictor than the participant-related variables that were considered by some distance. Of all the variables that were considered, then, formal distance is the single most important determinant of correct cognate guessing.

### **Partial formal distances**

The implementation of the Levenshtein algorithm that I used weighted all insertions, deletions and substitutions equally. However, previous studies suggested that some word parts could be more important than others in  $Lx$  item comprehension. First, written and spoken item comprehension may be more robust with respect to vowel insertions, deletions and substitutions than with respect to consonant operations. Second,  $Lx$  item comprehension may be more detrimentally affected by discrepancies between the stimulus and its known cognates at the beginning of the word than at the end of it. If consonants or word beginnings should indeed be weighted more heavily than in the Levenshtein implementation, one would expect to find the relevant partial Levenshtein distances to

make contributions to the models even when the overall Levenshtein distances have already been included.

To verify this prediction, I quantified the degree of consonantal and word-initial discrepancy between the  $Lx$  items and their German, English and French cognates in a fashion similar to Gooskens et al. (2008). The results, however, are not compellingly in favour of an account of consonantal or word-initial primacy: neither consonantal nor word-initial Levenshtein distances explain between-item variance in cognate guessing accuracy once overall Levenshtein distance is taken into consideration as well. In contrast to findings by Berthele (2011), Gooskens et al. (2008), Möller (2011), Möller and Zeevaert (2010) and Müller-Lancé (2003), the present data therefore suggest that consonants and word beginnings do not impact cognate guessing accuracy more than vowels and word middles and endings. Rather, the correlations between consonantal and word-initial Levenshtein distances and cognate guessing accuracy (see Table 9.3) seem to be by-products of their collinearity with the more important predictor, i.e. overall Levenshtein distance.

Note, however, that these results do not carry the implication that cognate guessing performance would not have been better *if* the participants had focussed more strongly on consonantal and word-initial similarities as opposed to vocalic and word-medial or -final similarities. What they do show is that participants do not necessarily do so without prior sensitisation or experience.

### Cognate frequency

On the basis of L1 findings, Van Heuven (2008) hypothesised that the frequency of occurrence of an  $Lx$  stimulus's L1 ( $L_2, \dots, L_n$ ) cognates is a determinant of the  $Lx$  stimulus's comprehension. Kürschner et al. (2008) tested this hypothesis empirically and found that cognate frequency was, in essence, an irrelevant predictor of spoken  $Lx$  word comprehension. The present study produced a similar result: the correlations between German, English and French cognate frequency and spoken word comprehension are low (Table 9.3) and the conditional variable importances of the frequency measures are effectively zero for the spoken stimuli. Cognate frequency does not seem to predict the comprehension of auditory  $Lx$  items.

Interestingly, however, cognate frequency does seem to be a predictor of written  $Lx$  item comprehension, as evidence both by the larger corre-

lation coefficients in Table 9.3 as by the non-zero conditional variable importances in Figure 9.5. Given that both German and English cognate frequency seemed to play a role in written *Lx* item comprehension, I created a new ‘Germanic’ frequency measure by averaging the German and English cognate frequencies for each word. This measure was associated with a respectable effect size in the mixed-effects model (ES:  $2.0 \pm 0.8$ , see Table 9.5).

The difference in importance of cognate frequency between the two modalities is congruent with differences in the role of crystallised resources in *Lx* cognate guessing that were uncovered in Part II. Crystallised resources such as L1 vocabulary knowledge seem to play a larger role in the written modality than in the spoken modality. Subjective word frequency is also a product of learning and exposure, i.e. it is a part of crystallised cognition. Jointly, both findings suggest that participants may find it easier to draw on stored linguistic knowledge in an *Lx* cognate guessing task when the stimuli are presented visually than when they are presented aurally.

### **Word length**

Stimulus length was a modest predictor of spoken word comprehension in Kürschner et al.’s (2008) study. The present study, by contrast, did not find any evidence suggesting that longer *Lx* stimuli are easier to understand: the bivariate correlation coefficients for the relationships between the by-item random intercepts and stimulus lengths are low and non-significant (Table 9.3) the conditional variable importances for stimulus length are effectively zero (Figures 9.5, 9.6 and 9.7). It can only be concluded that word length does not help to predict *Lx* comprehension in these data.

### **9.5.2 The use of multiple supplier languages**

In the written modality, bilingual (German–English) Levenshtein and frequency variables outperform their monolingual counterparts in terms of conditional variable importance. This is a further indication that participants in cognate guessing tasks do not solely rely on a related L1 but also on a related L2. Formal distance with respect to French and French cognate frequency, by contrast, do not seem to contribute to *Lx* item comprehension in the written modality.

Speculatively, this lack of importance of French-based variables may have multiple reasons. First, the participants' French skills may have been less developed than their English skills: to the extent that the participants' self-assessments are indeed valid (see Section 5.2.1 on page 70), these indicate that the participants' reading skills are on average somewhat lower in French ( $M = 3.0$ ,  $SD = 1.5$ ) than in English ( $M = 3.4$ ,  $SD = 1.3$ ; the CEFR levels were coded numerically from 1 (A1) to 6 (C2)). The participants may therefore have been less likely to rely on French as a supplier of transfer bases (Meißner and Burk, 2001; Williams and Hammarberg, 1998). Second, the number of written target words with a French cognate was limited: only ten written target words had a French cognate, with four of them having form-identical cognates in German or English. A higher proportion of target words with French cognates but without German or English cognates could have yielded different results. A third, related point is that the psychotypical distance (see Section 2.1.1 on page 18) from Swedish to German and English is likely to have been smaller than that to French, increasing the relative likelihood of German and English serving as supplier languages. It seems plausible, however, that this psychotypical distance can be influenced by including a higher proportion of target words related only to a French cognate. Thus, these results should not be interpreted as indicating that German-speaking Swiss participants *will not* under any circumstance draw on their knowledge of French when guessing cognates in a Germanic language, but rather that a task more conducive to French–L $x$  transfer is likely needed to detect such an effect.

In the spoken modality, only variables with respect to German seem to affect cognate guessing performance. In Section 8.2, I ventured the explanation that participants may be less able to draw on their crystallised resources in the spoken cognate guessing task than in the written one due to the time constraints associated with the spoken modality. This time constraint could especially hamper their efforts to engage in what Berthele (2008) called “linguistisches Probabilitätskalkül”. The present results on the item-related side are consistent with this explanation as they reveal that participants in spoken cognate guessing tasks are sensitive to linguistic information pertaining to the supplier language *par excellence*, i.e. L1 German, but not to other potential supplier languages. Presumably, inferences based on the L1 are more automatised than inferences based on the L2, ..., L $n$ , and given the

time constraints in the spoken modality, these other potential supplier languages could not have been taken sufficiently into account.

### 9.5.3 Postscript

The results regarding written  $Lx$  cognate guessing were wholly replicated in a study featuring 98 German-speaking Swiss participants and a total of 180 written Danish, Dutch, Frisian and Swedish words with German, English or French cognates (Vanhove and Berthele, forthcoming, b). In this study, too, Germanic Levenshtein distance and Germanic cognate frequency were independent predictors of  $Lx$  cognate guessing accuracy, whereas language-specific as well as partial Levenshtein distances did not contribute to the fit of the model. Additionally, the parameters of the two variables in a logistic mixed-effects model were highly similar to the ones reported in the present study: for overall Germanic Levenshtein distance, the regression parameter reported by Vanhove and Berthele (forthcoming, b) was  $-5.4 \pm 0.8$  compared to  $-5.2 \pm 1.2$  in the present study, and for log-transformed Germanic cognate frequency, it was  $0.34 \pm 0.09$  in Vanhove and Berthele (forthcoming, b) compared to  $0.33 \pm 0.12$  in the present study.

Vanhove and Berthele (forthcoming, b) recruited other participants and used different stimuli and  $Lxs$  than in the present study, yet the results of the two studies match closely. This speaks well for the robustness of the results respect to written cognate guessing that are reported here. For spoken cognate guessing, a similar parallel study has not yet been undertaken.

## Chapter 10

# Participant–item interactions

The analyses of the previous chapter showed which item-related characteristics co-determine cognate guessing accuracy. Although the findings are relevant and interesting in their own right, my principal goal was to find out to what extent the effects of these item-related characteristics change throughout the lifespan and as a function of cognitive variables that are themselves affected by ageing. In other words, what is of primary interest are the interactions between the item-related characteristics of the previous chapter and the participant-related variables of Part II. It is these interactions that I turn to in this chapter. For the sake of clarity, I focus on the cognitive variables that show the most diverging age trends, viz. fluid and crystallised intelligence, and leave the interactions with English proficiency and the number of foreign languages known by the participants out of consideration.

### 10.1 Possible interactions

The only item-related variables found to affect cognate guessing accuracy are the stimuli's overall formal overlap to known cognates and the frequency of these cognates. Based on both conceptual grounds and prior research, one can expect that the effects of these variables are not

the same for all participants but will instead vary as a function of the participants' fluid and crystallised resources. In what follows, I briefly sketch these conceptual grounds and previous findings.

### 10.1.1 Formal distance and fluid intelligence

By definition, participants with high fluid intelligence levels can deal relatively more flexibly with abstract patterns. To the extent that they can apply this ability in order to deal with obfuscated formal similarities between  $Lx$  stimuli and known cognates, it is to be expected not only that such participants outperform participants with lower fluid intelligence levels—a hypothesis that was only substantiated for the spoken modality—but also that formal discrepancies between  $Lx$  words and their  $L1, L2, \dots, Ln$  cognates do not affect them as much as the low-Gf participants. Thus, one would expect to find that the slope of the Levenshtein distance effect is less steep in participants who performed well on the Raven task.

### 10.1.2 Formal distance and crystallised intelligence

As I discussed in Section 2.1.2 on page 23, Berthele (2008) submitted that a larger linguistic repertoire may give rise to a greater degree of *Wahrnehmungstoleranz* in receptive multilingualism, i.e. greater flexibility in dealing with linguistic input that deviates from the own  $L1 (L2, \dots, Ln)$  norms (see also Teleman, 1981, as discussed in Section 3.1.3 on page 38). Greater flexibility with respect to  $Lx$ – $L1, L2, \dots, Ln$  discrepancies should be reflected in a weaker link between such discrepancies and  $Lx$  stimulus comprehension. Thus, to the extent that the size of the linguistic repertoire is associated with one's *Wahrnehmungstoleranz*, one would expect that participants with larger linguistic repertoires, as indicated by e.g. better scores on the crystallised intelligence task, experience a smaller overall Levenshtein distance effect (i.e. an effect with a gentler slope) than those with smaller linguistic repertoires.

### 10.1.3 Cognate frequency and crystallised intelligence

Other things being equal, the smaller one's vocabulary in a given language, the more strongly corpus frequency affects linguistic processing

in that language. For instance, participants with lower education levels show a steeper corpus frequency effect than do participants with higher education levels (Tainturier et al., 1992). As another example, bilinguals—who have less extensive vocabularies in each of their languages compared to monolinguals (e.g. Bialystok and Luk, 2012; Portocarrero et al., 2007)—likewise show stronger effects of corpus frequency on language processing than do monolinguals, and the frequency effect is stronger in their non-dominant than in their dominant language (e.g. Duyck et al., 2008; Gollan et al., 2008; Lemhöfer et al., 2008; Van Wijnendaele and Brysbaert, 2002). Diependaele et al. (2013) were able to show that this stronger word frequency effect in bilinguals is a by-product of their lower vocabulary knowledge levels in the target language rather than a direct result of their being bilingual.

Kuperman and Van Dyke (2013) argue that this frequency  $\times$  vocabulary size interaction does not need to indicate that participants with large vocabularies use frequency information in a different way. Rather, they argue, it may be a by-product of using objective, corpus-based frequency counts rather than subjective, participant-specific frequency estimates. Briefly, the driving force behind the frequency effect is not the stimuli's observed frequencies in a given corpus (say, SUBTLEX, CELEX or Google search results), but rather the number of times the stimuli have been previously encountered by the participants (i.e. subjective frequency); observed corpus frequencies merely serve as approximations of these subjective frequencies. However, as Kuperman and Van Dyke (2013) show, frequencies sampled from large corpora tend to be over-estimates of the subjective frequencies in the case of rare words, and even more so for participants who have had a relatively low degree of target language exposure. For highly frequent words, by contrast, the observed corpus frequencies are equally adequate approximations of the subjective frequencies for all participants. Since participants with large vocabularies tend to have had more extensive prior target language exposure (as likewise shown by Kuperman and Van Dyke, 2013), objective corpus-based frequencies are more accurate estimates of these participants' subjective frequencies than of those of participants with less exposure and smaller vocabularies. For these participants, the objective corpus-based frequencies should be adjusted increasingly more downwards for increasingly rarer words. The result is that the range of the subjective frequencies is larger than the range of the objective

frequencies, more so for participants with small vocabularies than for those with large vocabularies. This in turn creates the impression that the frequency effect is larger in participants with smaller vocabularies when objective frequency counts are used. When subjective frequency counts are used, however, this frequency  $\times$  vocabulary size interaction disappears. In sum, Kuperman and Van Dyke's (2013) analyses indicate that the frequency  $\times$  vocabulary size is not a genuine behavioural pattern but rather a by-product of using objective corpus-based frequencies.

The interactions observed between corpus-based frequency and vocabulary size prompt the question of whether such an interaction can also be found in the domain of *Lx* cognate guessing. In order to investigate this possibility, I make use of the SUBTLEX frequencies as the objective corpus frequency measures and the participants' WST scores as an indicator of their L1 vocabulary size. Unfortunately, subjective participant-specific frequency ratings of the stimuli's cognates are not available. This means that I cannot explore whether the corpus-based frequency  $\times$  vocabulary size interaction, if indeed found, can be attributed to a straightforward main effect of subjective frequencies.

To my knowledge, the interaction between fluid intelligence and word frequency has not been investigated nor do I see any reason why one might expect such an effect in cognate guessing. I therefore leave the interaction between fluid intelligence and cognate frequency out of consideration.

#### 10.1.4 Interactions with age

If cognate frequency and Levenshtein distance do indeed interact with fluid and crystallised intelligence in cognate guessing tasks, one would expect cognate frequency and Levenshtein distance to interact with age as well. First, young participants have lower crystallised intelligence levels than older participants and therefore could be expected to show a stronger frequency effect. Second, if higher fluid intelligence levels are associated with a lower degree of susceptibility to formal discrepancies in cognate guessing, one would expect participants around the 30 years of age mark, where fluid intelligence is at its peak, to show a smaller Levenshtein effect than both younger and older participants. Third, if higher crystallised intelligence likewise yields a lower dependence on formal overlap, one would expect the youngest participants to show the strongest Levenshtein effect. The second and third factor may conspire

so that the youngest participants (who generally have low Gf and Gc scores) show a very strong Levenshtein effect, the older participants (who have high Gc scores but fairly low Gf scores) a medium-strong Levenshtein effect and the 30-year-olds (who have both high Gc and Gf scores) the weakest Levenshtein effect.

## 10.2 Method of analysis

In order to explore the participant  $\times$  item interactions in the dataset, I make use of generalised additive models featuring random effects, i.e. GAMMs. As I wrote in Section 4.3.2 on page 63, GAMMs can model non-linear relationships between predictor variables and an outcome variable. In addition to modelling non-linear main effects, GAMMs can model non-linear interactions. This is typically done by means of *tensor product smooths*. The mathematical details are complex (see Wood, 2006, pp. 162–167) and need not concern us as end users; what is important is that tensor product smooths generalise the two-dimensional smooths (such as those plotted in Figure 6.2 on page 82) to higher dimensions. Interactions between continuous variables can in principle also be modelled in generalised linear mixed-effects models, but these assume linear main effects as well as linear interaction effects. GAMM-fitted tensor product smooths, by contrast, can reveal subtle non-linearities and thereby provide a more differentiated picture of the patterns in the data.

The GAMMs reported in this chapter were fitted using the `mgcv` package (version 1.7-24; Wood, 2013) for R (R Core Team, 2013). They were fitted separately for the written and for the spoken modality. The GAMMs featured (a) non-parametric main effects for the critical variables entering into the interactions that are under consideration, (b) tensor product interactions between these critical variables, (c) parametric main effects for the non-critical variables that have significant main effects in Tables 6.1 and 9.5 (written modality) or 6.2 and 9.7 (spoken modality) and (d) random intercepts for both participants and stimuli.

The main effects of the critical variables were approximately linear, which is why I opted to fit them in generalised linear mixed models in the previous chapters. Nevertheless, I deemed it preferable to allow these effects to be somewhat non-linear in this chapter so that the small non-linearities associated with the main effects would not be absorbed

by the tensor product smooths that model the interactions between the critical variables. This could have spuriously increased the interactions' significance. The tensor product smooths are fitted with cubic regression splines using `mgcv`'s `ti()` function.<sup>30</sup> The other significant but non-critical variables serve as control variables. Lastly, random intercepts for participants and stimuli are added to specify the data's dependency structure. Random slopes for the critical variables could not be added as these were not modelled parametrically (see Section 4.3.2). Apart from the non-linear interactions, the models in this chapter allow essentially the same conclusions to be drawn as those reported in Tables 9.5 and 9.7, but the latter models are superior for the purposes of discussing the main effects as they incorporate random slopes.

The `mgcv` package provides numerical information about the fitted non-linear terms, including their estimated degrees of freedom,  $\chi^2$  value and significance. These numerical estimates say little about the functional form of the non-linear terms, however, which is why these non-linear terms must be inspected visually. Non-linear interactions between two variables can be graphed in contour plots, which are two-dimensional representations of a three-dimensional surface on which points at the same 'height' (i.e. with the same fitted values) are connected by contour lines. Reading a contour plot of a non-linear interaction is thus essentially the same as reading a topographic map of hilly terrain.

### 10.3 Interactions between age and item-related predictors

In this section, I investigate the interaction between the item-related predictors and the participants' age. Then, in the next section, I turn to the interactions between the item-related variables and the participants' fluid and crystallised intelligence.

---

<sup>30</sup>Before version 1.7-23, `mgcv` only featured the `te()` function for fitting tensor product smooths, but according to the documentation of version 1.7-24, the `ti()` function is better suited to investigate main effects + interaction structures: using the `ti()` function produces a more reliable estimate of the interaction's contribution to the model when the main effects have already been taken into consideration.

### 10.3.1 Written items

The analyses in Chapter 9 showed that cognate guessing accuracy in the written modality is affected by the item-related variables of Germanic Levenshtein distance and Germanic cognate frequency. In order to explore whether these variables show varying effects as a function of age, I fitted their effects and their interactions with age non-linearly in a generalised additive model with crossed random intercepts for participants and items. The variable ‘ $\geq 1$  correct profile word translation’ had been found to be significantly associated with written cognate guessing accuracy in Chapter 6 and was therefore included as a control variable. Table 10.1 on the following page provides a numerical summary of this GAMM.

Table 10.1 suggests that the effect of Germanic cognate frequency may vary systematically as a function of age, but that the interaction between age and Germanic Levenshtein distance may not be statistically reliable.<sup>31</sup> Such numerical summaries reveal nothing about the functional form of the interactions, however, which is why I plotted the interaction between age and Germanic Levenshtein distance in the right-hand panel of Figure 10.1 and the one between age and Germanic cognate frequency in the right-hand panel of Figure 10.2. For expository purposes, the left-hand panel of each figure shows how a contour plot of the bare main effects (without an interaction) looks.

The left panels (without the interactions) permit the same inferences about the main effects as in Chapters 6 and 9. First, cognate guessing accuracy in the written modality increases fairly sharply throughout childhood and adolescence and develops slightly throughout adulthood: going from 10 to 25 years along the  $x$ -axes, the probability of a correct translation increases by about 1.5 log-odds, and going from 25 to 86 years, it shows a further improvement of about 0.5 log-odds. Second,

---

<sup>31</sup>Discussing results that are not significant at the 0.05 threshold may be anathema to some readers. However, these analyses are the first exploration of its kind of the interaction between participant- and item-related predictors in receptive multilingualism. Future studies may benefit from a description of the patterns observed in this study, even if they do not reach the traditional significance threshold: non-significant results are not by definition uninteresting. Furthermore, the  $p$ -values reported by `mgcv` are approximations that may be refined in future package versions. Consequently, one should not blindly rely on the significance tests, particularly if they yield  $p$ -values close to the 0.05 cut-off (on either side of it), but apply a healthy dose of researcher judgement in interpreting such data patterns.

**Table 10.1:** Generalised additive mixed-effect model modelling translation accuracy on written target words in function of non-linear interactions between age and item-related predictors. Panel (a): Parametric fixed effects, their standard errors and their significance. Panel (b): Smooth terms with their estimated degrees of freedom,  $\chi^2$ -statistics and significance. Panel (c): Modelled standard deviations of the random effects ( $\hat{\sigma}$ ). Parameter estimates are expressed in log-odds.

## (a) Parametric terms

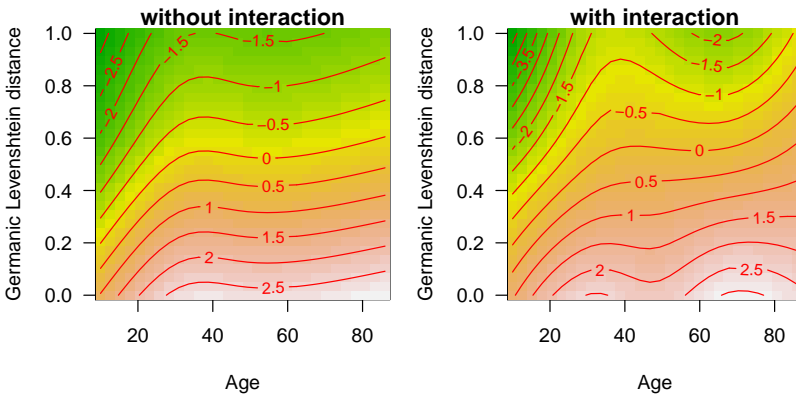
	Estimate $\pm$ SE	$p$
Intercept	$-0.83 \pm 0.22$	$<0.001$
$\geq 1$ correct profile word translation	$0.84 \pm 0.20$	$<0.001$

## (b) Smooth terms

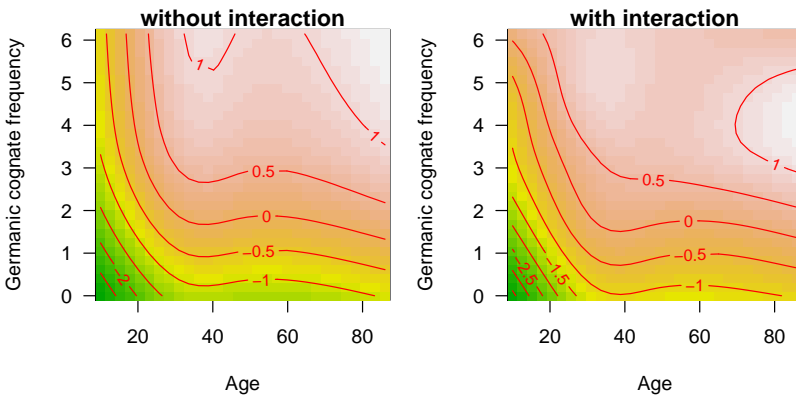
	Est. df	$\chi^2$	$p$
Age	3.4	79.3	$<0.001$
Germanic Levenshtein distance	1.2	18.8	$<0.001$
Germanic cognate frequency (log)	1.9	14.0	$<0.001$
Age $\times$ Germanic Levenshtein distance	5.7	13.5	0.077
Age $\times$ Germanic cognate frequency (log)	7.7	27.7	0.002

## (c) Random effects

	$\hat{\sigma}$
Random intercept by participant	0.81
Random intercept by items	1.4



**Figure 10.1:** GAMM-modelled effects of age and Germanic Levenshtein distance on cognate translation accuracy in the written modality. The probability estimates are in log-odds. Left: Without an interaction between age and Germanic Levenshtein distance. Right: With an interaction between age and Germanic Levenshtein distance.



**Figure 10.2:** GAMM-modelled effects of age and Germanic cognate frequency on cognate translation accuracy in the written modality. Left: Without an interaction between age and Germanic cognate frequency. Right: With an interaction between age and Germanic cognate frequency.

cognate guessing accuracy decreases as Germanic Levenshtein distance increases: going upwards along the  $y$ -axis in the left-hand panel of Figure 10.1, the probability of a correct translation drops by more than 4 log-odds.<sup>32</sup> Third, cognate guessing accuracy increases as Germanic cognate frequency increases: going upwards along the  $y$ -axis in the left-hand panel of Figure 10.2, the probability of a correct translation increases by roughly 2 log-odds.

In the right-hand panels, however, the size of these three effects depends on the position in the graph. In the right-hand panel of Figure 10.1, going upwards along the  $y$ -axis in the 10-to-20 year bracket, one sees the probability of a correct translation decrease from 1.0–1.5 to –4.5–3.5 log-odds, i.e. by 5.0 to 5.5 log-odds. Around age 40, by contrast, this probability decreases from less than 2.5 to nearly –1.5 log-odds, i.e. by less than 4.0 log-odds. In the older age brackets, the size of the Germanic Levenshtein distance effect varies between roughly 4.0 and 5.0 log-odds. Thus, young participants show a slightly stronger Levenshtein distance effect than older participants, and participants aged 30 to 50 show the weakest Levenshtein distance effect.

The interaction between age and Levenshtein distance seems to be driven mainly by the items with large Levenshtein distances. For items with Germanic Levenshtein distances up to about 0.4, the increase in cognate guessing accuracy between the ages of 10 and 30 (i.e. going right along the  $x$ -axis) is roughly 1.5 on the log-odds scale. For items with Germanic Levenshtein distances around 0.8, this increase is 2.5 log-odds, and for items with Germanic Levenshtein distances near 1.0, it is about 3.5 log-odds. It thus seems that the age-related increase in cognate guessing skills throughout childhood and adolescence is particularly pronounced when it comes to decoding stimuli with more obscured cognate relationships. The non-significance of the interaction term between age and Germanic Levenshtein distance serves as a warning against overinterpreting these effects, however.

The interaction between age and Germanic cognate frequency, depicted in the right-hand panel of Figure 10.2, seems to be more reliable statistically. In the youngest participants (10–20 years), the size of the Germanic cognate frequency effect is roughly 3.0 to 3.5 log-odds, whereas from roughly age 30 onwards, the change in probability along the  $y$ -axis

<sup>32</sup>This number is similar to, but different from, the one reported in Table 9.5, i.e.  $5.2 \pm 1.2$ , due to differences in the specification of the models.

varies between about 1.5 and 2.0 log-odds. Thus, young participants show a stronger frequency effect than older participants. Equivalently, one may say that the age-related development in cognate guessing skills between the age of 10 and 30 years is particularly pronounced in low-frequency stimuli.

### 10.3.2 Spoken items

The only item-related variable found to affect cognate guessing accuracy in the spoken modality was the stimuli's phonetic Levenshtein distance to their German cognates. In order to explore whether the effect of phonetic distance changes as a function of the participants' age, I fitted a GAMM with crossed random effects and a non-linear interaction between age and German Levenshtein distance. This model is presented in Table 10.2 on the next page, which shows that the non-linear interaction between age and German Levenshtein distance is statistically reliable.

For expository purposes, I again plotted the contour plot showing the non-linear interaction alongside a contour plot in which only the main effects were modelled in Figure 10.3 on page 165. The left-hand panel of Figure 10.3 allows the same inferences as those drawn in Chapters 6 and 9. First, cognate guessing accuracy in the spoken modality increases throughout childhood and young adulthood and then decreases again after reaching its peak in the 30-to-50 years bracket. Second, cognate guessing accuracy decreases by about 6.5 log-odds as German Levenshtein distance increases from 0 to 1.

The right-hand panel, however, reveals a subtle interaction between these two effects. For Levenshtein distances between 0.2 and 1.0, the size of the Levenshtein effect is about 5.0 log-odds for all participants. For Levenshtein distances between 0.0 and 0.2, however, the effect size is very small in participants aged between 10 and 25 (between about 0.0 and 0.5 log-odds). For participants aged around 40, the effect size is somewhat larger, namely slightly larger than 1.0 log-odds, but for participants aged 60 to 86, the effect size is between 2.0 and more than 2.5 log-odds. Thus, young participants seem to be hardly affected by small phonetic discrepancies between the stimuli and their German cognates, but such slight differences hamper the cognate guessing efforts of older participants considerably. Larger phonetic discrepancies, by contrast, seem to affect all participants indiscriminately.

**Table 10.2:** Generalised additive mixed-effect model modelling translation accuracy on spoken target words in function of non-linear interactions between age and item-related predictors. Panel (a): Parametric fixed effect, its standard error and its significance. Panel (b): Smooth terms with their estimated degrees of freedom,  $\chi^2$ -statistics and significance. Panel (c): Modelled standard deviations of the random effects ( $\hat{\sigma}$ ). Parameter estimates are expressed in log-odds.

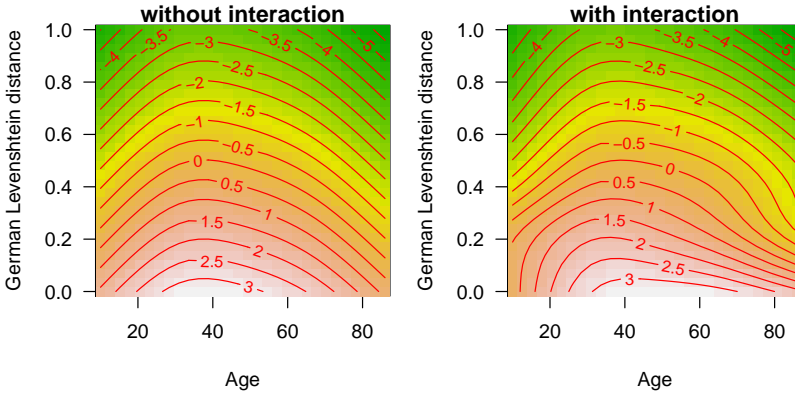
(a) Parametric terms			
	Estimate $\pm$ SE		$p$
Intercept	-1.0 $\pm$ 0.3		<0.001

(b) Smooth terms			
	Est. df	$\chi^2$	$p$
Age	3.5	96.0	<0.001
German Levenshtein distance	1.0	32.0	<0.001
Age $\times$ German Levenshtein distance	3.7	28.8	<0.001

(c) Random effects	
	$\hat{\sigma}$
Random intercept by participant	0.70
Random intercept by items	1.8



**Figure 10.3:** GAMM-modelled effects of age and German Levenshtein distance on cognate translation accuracy in the spoken modality. Left: Without an interaction between age and German Levenshtein distance. Right: With an interaction between age and German Levenshtein distance.

## 10.4 Interactions between cognitive and item-related predictors

### 10.4.1 Written items

The analyses in Section 10.3.1 indicated that younger participants show a stronger frequency effect than older participants and that participants aged 30 to 50 may show a weaker Levenshtein distance effect than the others (though the latter interaction was not statistically significant). In this section, I want to dig deeper into these age patterns by investigating the interactions between cognate frequency and Levenshtein distance on the one hand and the participants' fluid and crystallised intelligence on the other hand. To this end, I fitted a GAMM with crossed random intercepts for stimuli and participants and non-linear interactions between the participants' WST score and Germanic Levenshtein distance, between WST score and cognate frequency and between the participants' Raven score and Germanic Levenshtein distance. Since I had no a priori reasons to suspect an interaction between Raven score and cognate

frequency, I did not include such an interaction in the model. The model also included three control variables that were found to affect cognate guessing accuracy in Chapter 7, viz. ‘ $\geq 1$  correct profile word translation’, the number of foreign languages in the participants’ repertoire and their performance on the English tests. The GAMM is presented in Table 10.3 on the next page.

I first turn to the interaction between the participants’ WST score—the crystallised intelligence indicator—and Germanic Levenshtein distance. This interaction is presented in Figure 10.4 on page 168 alongside a contour plot of the bare main effects. The left-hand panel of Figure 10.4 allows the same inferences as those discussed in Chapters 7 and 9 and will not further be discussed. The right-hand panel suggests that participants with low WST scores (e.g. 10) show a Levenshtein distance effect of about 6.5 log-odds, whereas this effect is smaller in participants with average WST scores (e.g. 5.0 log-odds for participants with a WST score of 25) and smaller still in participants with high WST scores (e.g. 4.0 log-odds for participants with a WST score of 35). This interaction seems to be largely driven by the items showing large orthographic differences towards their German or English cognates: whereas the size of the WST effect varies roughly between 2.0 and 2.5 log-odds for items with a Germanic Levenshtein distance of less than 0.6, it measures about 3.5 log-odds for items with a Levenshtein distance of 0.7 and roughly 5.0 log-odds for items with a Levenshtein distance of 0.9. Thus, high crystallised intelligence levels may be particularly useful for decoding stimuli with no or highly obscured cognate relationships to Germanic translation equivalents. However, these results must be taken with a grain of salt given that the interaction is not statistically significant (see Table 10.3).

Second, Figure 10.5 on page 169 shows the interaction between the participants’ Raven score—the fluid intelligence indicator—and the Germanic Levenshtein distance variable. The main effect of Raven score was small in size and not significant as shown in the left-hand panel. The right-hand panel, however, suggests that there may be a cross-over interaction between Raven score and Levenshtein distance: for items with low Levenshtein values (up to about 0.2), cognate guessing accuracy actually *decreases* with increasing Raven task performance, if only slightly. For items with higher Levenshtein values, by contrast, cognate guessing accuracy improves somewhat with increasing Raven

**Table 10.3:** Generalised additive mixed-effect model modelling translation accuracy on written target words in function of non-linear interactions between participant- and item-related predictors. Panel (a): Parametric fixed effects, their standard errors and their significance. Panel (b): Smooth terms with their estimated degrees of freedom,  $\chi^2$ -statistics and significance. Panel (c): Modelled standard deviations of the random effects ( $\hat{\sigma}$ ). Parameter estimates are expressed in log-odds.

(a) Parametric terms

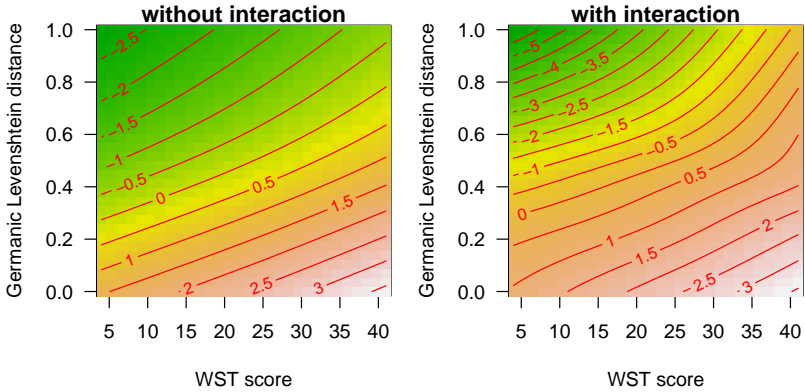
	Estimate $\pm$ SE	<i>p</i>
Intercept	-0.81 $\pm$ 0.22	<0.001
$\geq 1$ correct profile word translation	0.51 $\pm$ 0.16	0.002
Number of foreign languages	0.15 $\pm$ 0.07	0.031
English proficiency	0.13 $\pm$ 0.05	0.011

(b) Smooth terms

	Est. df	$\chi^2$	<i>p</i>
WST score	1.4	41.3	<0.001
Raven score	1.0	1.1	0.305
Germanic Levenshtein distance	1.0	19.1	<0.001
Germanic cognate frequency (log)	1.9	14.6	<0.001
WST score $\times$ Germanic Levenshtein distance	3.3	8.4	0.109
Raven score $\times$ Germanic Levenshtein distance	1.0	10.4	0.001
WST score $\times$ Germanic cognate frequency (log)	2.4	24.5	<0.001

(c) Random effects

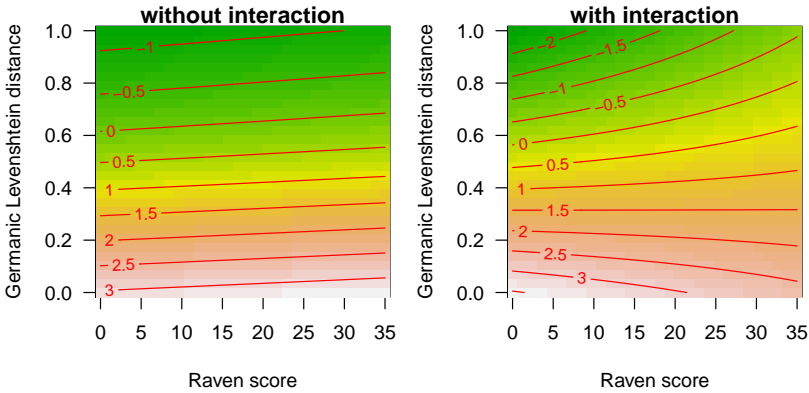
	$\hat{\sigma}$
Random intercept by participant	0.65
Random intercept by items	1.4



**Figure 10.4:** GAMM-modelled effects of crystallised intelligence (WST score) and Germanic Levenshtein distance on cognate translation accuracy in the written modality. The probability estimates are in log-odds. Left: Without an interaction between WST score and Germanic Levenshtein distance. Right: With an interaction between WST score and Germanic Levenshtein distance.

scores. Particularly for items with Levenshtein values of 0.8 and higher, the Raven effect is noticeable with an effect size of about 1.5 log-odds. This cross-over interaction is also reflected in the systematically varying effect sizes of the Levenshtein distance effect for different Raven score levels: for the lowest Raven score levels, the Levenshtein effect size measures more than 5 log-odds; for the highest Raven score levels, it measures only about 3 log-odds. In sum, high fluid intelligence levels are particularly advantageous when decoding stimuli showing little orthographic overlap with their German or English cognates, but may actually be disadvantageous, if only slightly, when decoding stimuli that are orthographically highly similar to their Germanic cognates. This interaction appears to be statistically reliable (see Table 10.3).

The last interaction—the one between the participants’ WST score and the items’ Germanic cognate frequency—is plotted in Figure 10.6 on page 170. The right-hand panel of Figure 10.6 shows that the frequency effect is about 3.0 to 3.5 log-odds for participants with WST scores up to about 25. In participants with WST scores of 35 and

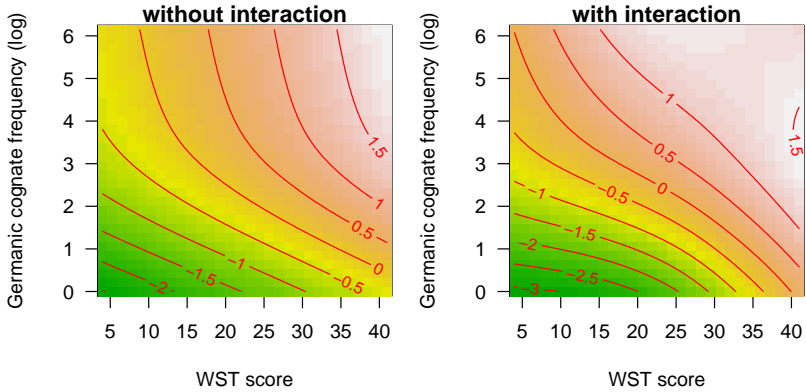


**Figure 10.5:** GAMM-modelled effects of fluid intelligence (Raven score) and Germanic Levenshtein distance on cognate translation accuracy in the written modality. The probability estimates are in log-odds. Left: Without an interaction between Raven score and Germanic Levenshtein distance. Right: With an interaction between Raven score and Germanic Levenshtein distance.

higher, the frequency effect measures only between 1.5 and 2.0 log-odds. From another perspective, it can be seen that the WST effect is particularly strong for low-frequency items (e.g. 3.0 log-odds for items with a Germanic cognate frequency of 0 log-units), but that it diminishes in strength as cognate frequency increases: for items with a Germanic cognate frequency of 4 log-units, the WST effect is about 2.0 log-odds; for items with a cognate frequency of 6 log-units, it is less than 1.5 log-odds. Thus, cognate frequency is a less important predictor of cognate guessing accuracy in participants with high crystallised intelligence levels and, conversely, crystallised intelligence levels are less important when the stimuli have high-frequency cognates. This interaction appears to be statistically reliable (see Table 10.3).

### 10.4.2 Spoken items

The analyses in Section 10.3.2 suggest that slight phonetic discrepancies between the stimuli and their German cognates do not affect cognate guessing accuracy in young participants as severely as in older partic-



**Figure 10.6:** GAMM-modelled effects of L1 vocabulary knowledge (WST score) and Germanic cognate frequency on cognate translation accuracy in the written modality. The probability estimates are in log-odds. Left: Without an interaction between WST score and Germanic cognate frequency. Right: With an interaction between WST score and Germanic cognate frequency.

ipants. In order to investigate whether this interaction can be traced back to interactions between phonetic Levenshtein distance and crystallised intelligence and between phonetic Levenshtein distance and fluid intelligence, I fitted a GAMM with crossed random intercepts for stimuli and participants as well as non-linear interactions between Levenshtein distance and WST score and between Levenshtein distance and Raven score. The variables English proficiency and backward digit span were found to be related to cognate guessing accuracy in the spoken modality in Chapter 7 and were included in the model as control variables. Table 10.4 on the next page presents the fitted model.

The inferences regarding the main effects are the same as those discussed in Chapters 6 and 7 and will not further be discussed here. Instead, I first turn to the interaction between German Levenshtein distance and the participants' WST score—the crystallised intelligence indicator, which is presented in the right-hand panel of Figure 10.7 on page 172. The contour plot suggests that participants with low to average WST scores show a smaller effect of Levenshtein distance

**Table 10.4:** Generalised additive mixed-effect model modelling translation accuracy on spoken target words in function of non-linear interactions between participant- and item-related predictors. Panel (a): Parametric fixed effects, their standard errors and their significance. Panel (b): Smooth terms with their estimated degrees of freedom,  $\chi^2$ -statistics and significance. Panel (c): Modelled standard deviations of the random effects ( $\hat{\sigma}$ ). Parameter estimates are expressed in log-odds.

(a) Parametric terms

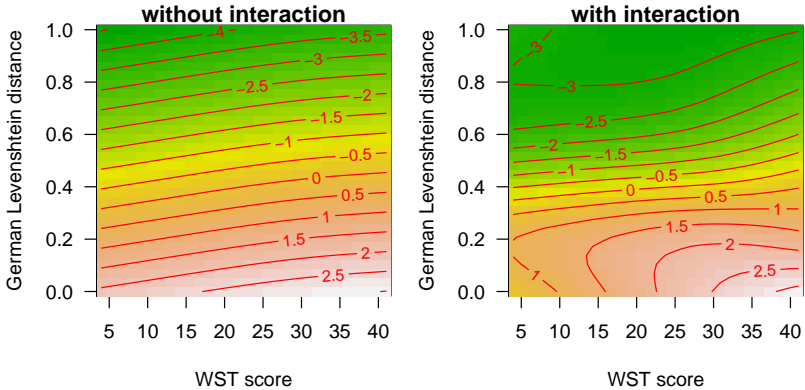
	Estimate $\pm$ SE	$p$
Intercept	-1.0 $\pm$ 0.3	<0.001
English proficiency	0.20 $\pm$ 0.05	<0.001
Backward digit span	-0.075 $\pm$ 0.039	0.053

(b) Smooth terms

	Est. df	$\chi^2$	$p$
WST score	1.0	5.1	0.024
Raven score	1.0	20.4	<0.001
German Levenshtein distance	2.3	35.9	<0.001
WST score $\times$ German Levenshtein distance	5.5	13.1	0.085
Raven score $\times$ German Levenshtein distance	6.0	16.3	0.041

(c) Random effects

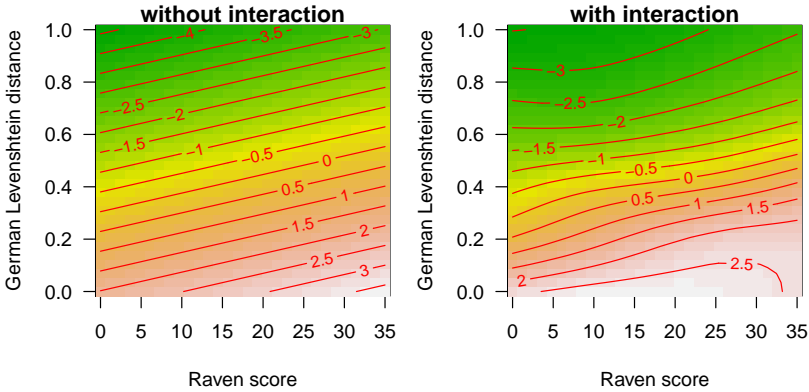
	$\hat{\sigma}$
Random intercept by participant	0.64
Random intercept by items	1.7



**Figure 10.7:** GAMM-modelled effects of crystallised intelligence (WST score) and German Levenshtein distance on cognate translation accuracy in the spoken modality. The probability estimates are in log-odds. Left: Without an interaction between WST score and German Levenshtein distance. Right: With an interaction between WST score and German Levenshtein distance.

than participants with WST scores of 30 and higher—contrary to the expectations set out in Section 10.1. The interaction seems to be driven primarily by the items with low Levenshtein distance values: participants with WST levels of about 25 and lower hardly show a Levenshtein effect for items with Levenshtein values between 0 and 0.2; participants with WST levels of 30 and higher, by contrast, do show a Levenshtein effect for these stimuli. This interaction is not significant, however (see Table 10.4).

The second interaction, the between German Levenshtein distance and Raven score, is plotted in the right-hand panel of Figure 10.8. The contour plot suggests that the overall effect of Levenshtein distance is roughly stable across the range of the Raven variables, hovering around 5.5 log-odds. However, a subtle interaction may be present in that the Raven score effect may be somewhat different in different Levenshtein distance brackets. For the items showing near-complete overlap with their German cognates (Levenshtein distances lower of 0.1 and lower), the size of the Raven effect is roughly 0.5 log-odds. For items with Levenshtein



**Figure 10.8:** GAMM-modelled effects of fluid intelligence (Raven score) and German Levenshtein distance on cognate translation accuracy in the spoken modality. The probability estimates are in log-odds. Left: Without an interaction between Raven estimates and German Levenshtein distance. Right: With an interaction between Raven score and German Levenshtein distance.

distances between roughly 0.2 and 0.5, the effect is appreciably larger at more than 1.5 log-odds. For items with even larger Levenshtein distances, the Raven effect measures only about 1 log-odds. Thus, fluid intelligence may be particularly useful when decoding spoken stimuli that show a modest degree of discrepancy towards their German cognates. As a caution against overeager post-hoc interpretations of this subtle interaction, however, I point out that it is not very robust statistically: while its significance level is slightly under the traditional 0.05 threshold (see Table 10.4), removing the non-significant interaction between WST score and German Levenshtein distance from the model puts it above the cut-off mark.

## 10.5 Discussion

### 10.5.1 Written items

The analyses in Section 10.3.1 suggest that the effect of Germanic Levenshtein distance on cognate guessing accuracy in the written modality is stronger in children and adolescents than in 30-to-50-year-olds, with older participants showing an in-between Levenshtein effect. This effect is particularly pronounced in stimuli with Levenshtein values higher than about 0.50. Additionally, young participants show a stronger effect of Germanic cognate frequency than participants aged 30 years and older such that there is hardly any age effect for stimuli with highly frequent cognates in German and English, whereas the age effect is very strong for stimuli that have low-frequency (or no) cognates in German and English. Both interactions are largely in line with the expectations outlined in Section 10.1.4, although the age  $\times$  Levenshtein interaction may not be statistically reliable.

If age interacts with item-related variables, then it may be the case that interactions between item-related variables and the age-labile and age-stable facets of cognition, i.e. fluid and crystallised intelligence, underlie this interplay. As I argued in Section 10.1, both high fluid and crystallised intelligence levels could be expected to be associated with a weaker effect of Levenshtein distance. Likewise, high crystallised intelligence levels could be expected to be associated with a weaker role of cognate frequency.

With respect to the interaction between cognate frequency and crystallised intelligence, this prediction was fully borne out: participants with high WST scores showed smaller frequency effects than participants with low WST scores. It thus seems that previous findings demonstrating an interaction between (objective) word frequency and lexical processing in known languages carry over to *Lx* cognate guessing tasks as well, at least in the written modality. WST scores increase dramatically throughout childhood and adolescence and remain largely stable throughout the rest of the lifespan. This age-patterning is reflected in the age  $\times$  frequency interaction. Whether this is due to the reason put forward by Kuperman and Van Dyke (2013), i.e. that differences in subjective frequencies cause such a spurious interaction between vocabulary knowledge and

objective frequencies, is a question that could not be answered in this study, however.

The results with respect to the interactions with Levenshtein distance are less clear-cut. First, the direction of the interaction between crystallised intelligence and Germanic Levenshtein distance was in line with what was expected: participants with high WST scores show a smaller Levenshtein effect than those with low WST scores. This effect is particularly pronounced in stimuli with Levenshtein values of 0.50 and more. This would suggest that participants with large linguistic repertoires show a larger degree of *Wahrnehmungstoleranz* and are less hampered by formal differences when guessing the meaning of written Lx cognates. Given the age-patterning of crystallised intelligence, it would also go some way in explaining why the youngest participants, whose crystallised intelligence levels are well below those of the other participants, seem to be most sensitive to formal differences. However, this interaction, while in line with the expectations, may not be statistically reliable.

Second, participants with high fluid intelligence levels showed a weaker effect of Germanic Levenshtein distance than did participants with low fluid intelligence levels. This interaction squares with the expectations and, given the age-patterning of fluid intelligence, may likewise help to explain why young participants show the strongest Levenshtein distance effect: they have both low fluid and crystallised intelligence levels, and both would seem to yield a stronger reliance on formal similarities. Moreover, older participants have somewhat stronger Levenshtein effects than 30- to 50-year-olds. Since the former have low fluid intelligence scores but high WST scores and the latter have relatively high scores in both domain, this age  $\times$  Levenshtein distance interaction can similarly be explained in terms of cognitive factors.

However, the fluid intelligence  $\times$  Levenshtein distance interaction revealed an intriguing and unexpected cross-over in the effect of the Raven variable. For stimuli with high Levenshtein values, the effect was in the expected direction: better Raven scores were associated with higher cognate guessing accuracy. For stimuli with low Levenshtein values, however, higher Raven scores were associated with slightly *lower* cognate guessing accuracy. This cross-over interaction may explain why no main effect of fluid intelligence was found in the written modality (see Chapter 7) as the negative effect cancels the positive trend partially

out. It does raise the question of why higher Raven scores are associated with slightly lower cognate guessing performance for items with low Levenshtein values, however. One possible explanation goes as follows. Participants with high fluid intelligence levels are arguably more adept at coping with obscured cognate relationships as they can treat the stimuli's forms more flexibly by abstracting away from  $Lx-L1$ ,  $L2$ , ...,  $L_n$  differences and instead establishing similarities. When cognate relationships are obscured, this is usually advantageous, but when they are not, this flexibility might backfire. For example, showing some flexibility with regard to formal discrepancies is necessary when decoding the stimulus *förutsättning* (Gm. *Voraussetzung* 'requirement'). When decoding a stimulus like *hård* 'hard', not much flexibility is required; in fact, decoders who are too flexible can sometimes come up with incorrect translations such as *Herz* 'heart' or *Herd* 'stove'. That said, the negative effect of fluid intelligence is relatively small.

### 10.5.2 Spoken items

Only one item-related variable was found to affect cognate guessing accuracy in the spoken modality: the stimuli's phonetic Levenshtein distance towards their German cognates. I therefore only investigated how ageing affects the strength of the effect of this variable. Contrary to my expectations, the youngest participants showed the weakest Levenshtein effect and the oldest showed the strongest one. The effect was primarily driven by stimuli with low Levenshtein distances: older participants performed essentially at ceiling on form-identical cognates, but their performance dropped steeply if the Swedish stimuli showed even minute differences towards their German counterparts. Young participants, by contrast, started at a lower baseline, but their performance was hardly affected by tiny formal differences in the cognate relationships.

An exploration of the interactions between crystallised and fluid intelligence on the one hand and Levenshtein distance on the other did not yield a satisfactory account of this unexpected age  $\times$  Levenshtein interaction. First, participants with low crystallised intelligence levels seemed to be more sensitive to formal discrepancies between the stimuli and their German counterparts. The direction of this interaction is opposite to both the prior expectations and the findings for the written modality. Furthermore, it is not statistically robust. Second, participants with high fluid intelligence levels may be less sensitive to small phonetic

Levenshtein distances than participants with low fluid intelligence levels. Since fluid intelligence is at its peak around age 30 in the present participant sample, however, one would expect 30-year-olds to show the lowest degree of sensitivity to small formal differences—not children and adolescents. Additionally, the fluid intelligence  $\times$  Levenshtein interaction may not be statistically robust either.

In sum, the age  $\times$  Levenshtein interaction in the spoken modality cannot satisfactorily be explained in terms of the cognitive factors considered in this study. To the best of my knowledge, research in related fields does unfortunately not offer much in the way of an explanation either. For instance, decoding *Lx* stimuli showing small phonetic discrepancies from their L1 cognates may be akin to listening to accented L1 speech. Thus, if older adults have greater difficulties listening to accented L1 speech than younger adults, a link to research on accent perception could be established. However, in a review on accent perception across the lifespan, Cristia et al. (2012) conclude that

while adults do have greater difficulty with accented than unaccented speech, the size of this effect is not significantly greater for older than younger listeners (p. 8).

In the absence of a convincing explanation of why older adults seem to be more strongly affected by mild phonetic *Lx*–L1 differences than younger adults and children, whereas they are equally strongly affected by larger discrepancies, I am inclined to offer this as a tentative finding that is in need of both explanation and replication.



## Part IV

# Conclusions



# Chapter 11

## Synthesis and new directions

### 11.1 Synthesis

This thesis investigated the lifespan development of a key skill in foreign language learning and receptive multilingualism: the ability to make sense of isolated written and spoken words in an unknown language but with cognates in known languages. Of specific interest was the question of how such age-related developments in this skill could be attributed to cognitive and linguistic factors. The data from a Swedish cognate guessing task administered to a cross-sectional sample of multilingual Swiss-German participants aged 10 to 86 years were analysed from three vantage points: first with respect to the inter-individual differences in cognate guessing performance, second with respect to between-item differences in overall cognate guessing accuracy and lastly focussing on the interplay between participant- and item-related characteristics. The results of these analyses are discussed in detail in Chapters 8, 9 and 10 and will not be rehashed in full. Here, I will knit together the main findings emerging from these different analyses. The applied implications of this project will not be taken up here but are discussed by Berthele and Vanhove (forthcoming).

The analyses in Part II show that cognate guessing skills develop differently in the written and in the spoken modality. In both modalities, cognate guessing performance increases sharply throughout childhood and adolescence. Performance keeps improving slightly throughout the adult lifespan in the written modality, but in the spoken modality, performance plummets starting around age 50. The cause for these diverging age trends seems to be that the visual mode of presentation enables cognate guessers to draw on their crystallised (knowledge-based) resources to a greater extent than the aural mode. In the aural mode, cognitive flexibility (in the form of fluid intelligence) takes on greater importance.

The item-based results of Chapter 9 in Part III complement these findings as they show that participants are sensitive almost exclusively to formal similarities and discrepancies towards the most natural supplier language, i.e. L1 German, in the spoken modality. In the written modality, the same participants are also sensitive to similarities and discrepancies vis-à-vis a related foreign language, English, as well as to frequency information pertaining to German and English. This suggests that information of different kinds (form and frequency) and from different sources (German and English) can more efficiently be integrated when guessing the meaning of written cognates. I have speculated that time pressure differences between the two modes of presentation are the main contributor to these by-modality differences, but further research will be needed to test this hypothesis (see Section 11.2).

Lastly, the joint consideration of item- and participant-related effects and their interactions yielded a differentiated picture of the interplay between both kinds of variables in cognate guessing. In the written modality, an improvement in cognate guessing skills throughout childhood and adolescence is noticeable for all kinds of items, but the increase is especially pronounced for ‘difficult’ stimuli—those with high formal distances towards their German and English cognates and with low cognate frequencies (see Figure 10.1 on page 161 and Figure 10.2 on page 161). This age trend seems to be governed by the participants’ cognitive development: fluid and crystallised resources improve into young adulthood and these cognitive improvements allow the participants to rely less and less on formal similarities between the stimuli and their cognates. Thus, objective similarity becomes increasingly less important

as a determinant of perceived similarity as the participants develop cognitively (for the difference between objective and perceived similarity, see Section 2.1.1 on page 18). At the same time, as one's crystallised resources develop, so does one's ability to access rarer linguistic items to serve as a basis for guessing the meaning of unknown words.

In the spoken modality, the picture that emerges is less clear. Participants in middle through older age seem to become increasingly sensitive to small phonetic differences between the stimuli and their German counterparts. Younger participants are largely insensitive to such minute discrepancies (see Figure 10.3 on page 165). It thus seems that the age-related increase in spoken cognate guessing ability throughout younger age and the age-related decrease throughout older age are not each other's perfect mirror image: the increase is partly associated with increases in cognate guessing accuracy for the easiest items, i.e. form-identical cognates, whereas the decrease is partly due to the increased sensitivity to small phonetic differences; baseline performance for the form-identical cognates only drops slightly in the last decades. As of yet, however, it is not known which cognitive changes underlie these age differences.

In summation, the ability to make sense of words in an unknown language with known cognates shows an age-related development that is modality-dependent. Cognitive changes underlie this development to a substantial extent. Furthermore, age interacts with item-related effects so that the relationship between the objective similarity between unknown words and known cognates and cognate guessing accuracy is not constant across the lifespan.

## 11.2 Avenues for further research

A key finding of the present study is that it is easier to bring to bear crystallised resources when guessing the meaning of written cognates compared to spoken cognates. At present, I suspect that it is the short-lived nature of aural presentation that is the cause for these by-modality differences. This hypothesis can be tested in an experimental design with two (within-subject) conditions: in one condition, participants are presented with written cognates that remain on-screen until they venture their final guess (as in the present study); in the other, the written stimuli disappear after the participants have had time to read it once. Alternatively, the first condition could feature spoken cognates

that are presented only once, whereas the participants could be given the option to replay the stimuli several times in the second condition. The prediction is that crystallised and fluid resources interact with the experimental condition: longer or repeated exposure will lead to a higher involvement of crystallised resources and a lower involvement of fluid intelligence. Additionally, longer exposure will lead to more guesses that are not only based on the L1.

Furthermore, two results in the present study were difficult to explain. First, working memory capacity, as measured using a backward digit span task, was associated negatively with cognate guessing accuracy in the spoken modality (Chapter 7). Second, also in the spoken modality, the youngest participants were hardly affected by small formal  $Lx$ -L1 discrepancies whereas the oldest were strongly affected by such minute differences. This finding ran contrary to the expectations and could not satisfactorily be attributed to the effects of fluid or crystallised intelligence (Chapter 10). Further work is needed to verify whether these two puzzling findings are in fact empirically robust and, if so, to provide coherent explanations for them.

Such future studies could also improve on the design of the present study. In addition to including more items per modality than was possible in the present study, they could extract more accurate assessments of the participant-related constructs that were considered. For reasons of time, each cognitive construct could only be assessed once. In an ideal world, working memory capacity as well as fluid and crystallised intelligence would be tested by means of multiple tests in a latent variable approach as recommended by Conway et al. (2005). In addition, the participants' foreign language vocabulary, especially in English and French, could be subjected to a more targeted assessment. The LexTALE tests (Brysbaert, 2013; Lemhöfer and Broersma, 2012), which were published during the course of the present project, would be a relatively time-friendly means of accomplishing this.

Lastly, cognate guessing tasks are admittedly highly reductionistic and cannot capture the full complexity of receptive multilingualism. Future studies may want to extend their scope to the comprehension of  $Lx$  phrases, sentences and texts. Cognate guessing will still be an integral part of making sense of such larger chunks, but its effects will be probabilistic and will be modulated by other cues. As a simplified example, consider the Danish word *hvid*. In the absence of context, a

German-speaking cognate guesser with knowledge of English may estimate that this word likely means *weit* ‘wide’ and may judge *weiß* ‘white’ and alternatives to be considerably less likely meanings. In a cognate guessing task, the participant would thus be more likely to answer *weit*. When presented with the noun phrase *det hvide hus*, however, the participant may realise that *det* is an article and that *hus* probably means *Haus* ‘house’. The probability of encountering the trigram *das weiße Haus* ‘the white house’ is vastly higher than that of encountering the trigram *das weite Haus* ‘the wide house’ (according to Google Ngram Viewer, available from <http://books.google.com/ngrams>), so that *weiß* may yet become the preferred reading of *hvid(e)* given the (minimal) context. In a more elaborate context featuring references to *Washington D.C.*, *Jimmy Carter* or *den amerikanske præsident*, the reading *weit* may become so unlikely that it does not occur.

This proposed probabilistic mechanism (inspired by the Bayesian models of L1 reading and speech recognition developed by Norris, 2006, and Norris and McQueen, 2008) still assigns an important role to the ability to make accurate cognate guesses, but it allows the guesses for individual words to be adjusted in function of how certain the participant is of other inferences (including other cognate guesses). Participants likely differ considerably in how much they can or are willing to update their beliefs in the face of new contextual evidence. Furthermore, different contextual cues may be picked up by different participants—if you do not know who Jimmy Carter is, seeing his name is unlikely to shift your beliefs about the meaning of *hvide*. Thus, the challenge for future research will be to isolate the factors that modulate the effects of cognate guessing in receptive multilingualism. And I believe that the most fruitful way to go about this is to consider the influence of phrasal context next.



# Appendices



# Appendix A

## Items for the cognate guessing task

### A.1 Written items

**Table A.1:** Written stimuli used in the Swedish cognate guessing task along with their English translations and German, English and French cognates.

Stimulus	Translation	German	English	French
alltid	always	allzeit		
avskaffa	to abolish	abschaffen		
bakgrund	background		background	
behärska	to master	beherrschen		
borgmästare	mayor	Bürgermeister		
byrå	bureau	Büro	bureau	bureau
bäbis <sup>33</sup>	baby	Baby	baby	bébé
cyckel <sup>34</sup>	(bi)cycle	Zyklus	cycle	cycle

(Continued on next page)

<sup>33</sup> *Bäbis* is actually a common misspelling for *bebis*. This misspelling is wholly inconsequential for the present purposes.

<sup>34</sup> *Cyckel* is a misspelling for *cykel* (which in fact is pronounced as though it were written *cyckel*). Again, this misspelling is inconsequential for the present purposes.

Table A.1 – continued from previous page

Stimulus	Translation	German	English	French
fiende	enemy	Feind	fiend	
fåtölj	fauteuil	Fauteuil	fauteuil	fauteuil
försiktig	careful	vorsichtig		
förutsättning	requirement	Voraussetzung		
full	full	voll	full	
hård	hard	hart	hard	
kanel	cinnamon			cannelle
kejsar	emperor	Kaiser		
kniv	knife		knife	
kung	king	König	king	
kyrka	church	Kirche	church	
kyssa	to kiss	küssen	kiss	
löpa	to run	laufen		
mjölk	milk	Milch	milk	
möjlig	possible	möglich		
rytmisk	rhythmic	rhythmisch	rhythmic	rythmique
rådhus	town hall	Rathaus		
saliv	saliva		saliva	salive
skola	school	Schule	school	école
skrubba	to scrub	schrubben	scrub	
skyskrapa	skyscraper		skyscraper	
sitta	to sit	sitzen	sit	
skön	beautiful	schön		
spegel	mirror	Spiegel		
språk	language	Sprache		
stjärn <sup>35</sup>	star	Stern	star	star
söka	to search	suchen	seek	
torsdag	Thursday	Donnerstag	Thursday	
tunga	tongue	Zunge	tongue	
tvivla	to doubt	zweifeln		
tårta	cake	Torte	tart	tarte
varm	warm	warm	warm	

(Continued on next page)

<sup>35</sup>While *stjärn* is an existing Swedish word meaning ‘blaze, star’ (i.e. a white spot on a dark horse), the cognate referring to the celestial object or to famous people is actually *stjärna*.

**Table A.1 – continued from previous page**

Stimulus	Translation	German	English	French
viktig	important	wichtig		
värld	world	Welt	world	
ytterst	extreme	äusserst		
öppna	to open	öffnen	open	
översätta	to translate	übersetzen		
barn	child	(profile word)		
häst	horse	(profile word)		
leka	to play	(profile word)		
mycket	very, much	(profile word)		
städa	to clean	(profile word)		

## A.2 Spoken items

**Table A.2:** Spoken stimuli used in the Swedish cognate guessing task along with broad IPA transcriptions, English translations and German, English and French cognates. Phonological consonant length in Swedish words was not transcribed.

Stimulus	Transcription	Translation	German	English	French
bliva	'blivva	to stay	bleiben		
blomma	'blomma	flower	Blume	blom	
bränna	'brænna	to burn	brennen	burn	
butelj	bʉ'telj	bottle		bottle	bouteille
choklad	ʃu'klad	chocolate	Schokolade	chocolate	chocolat
egenskap	'e:gen,ska:p	characteristic	Eigenschaft		
elev	e'lev	pupil			élève
ensam	'en,sam	lonely	einsam		
fotboll	'fʉt,bol	football	Fussball	football	football
fråga	'fro:ga	question; to ask	Frage (fragen)		
fräsch	fr:e:f	fresh	frisch	fresh	
fönster	'fön:ster	window	Fenster		fenêtre
försöka	fö:'söka	to try	versuchen		
först	fö:st	first		first	
försvinna	fö:'sʃvina	to disappear	verschwinden		
grupp	gr:ep	group	Gruppe	group	groupe
gå	go:	to go	gehen	go	
hemlig	'hæmlig	secret	heimlich		
ingenjör	in:ʃen,'jœr	engineer	Ingenieur	engineer	ingénieur
intryck	'in,tryk	impression	Eindruck		
is	i:s	ice	Eis	ice	

(Continued on next page)

Table A.2 – continued from previous page

Stimulus	Transcription	Translation	German	English	French
konst	konst	art	Kunst		
korruption	kor'p'ʃju:m	corruption	Korruption	corruption	corruption
kägel	'ʃe:ʒel	cone	Kegel		
kärnkraftverk	'ʃæ:ŋkra:ft,'vɛ:rk	nuclear power station	Kernkraftwerk		
larm	lar:m	noise; alarm	Lärm (Alarm)	alarm	alarme
lång	lɔ:ŋ	long	lang	long	long
märkvärdig	'mæ:rk,'væ:dɪg	remarkable	merkwürdig		
nackdel	'nɑ:k,dɛ:l	disadvantage	Nachteil		
paraply	para'pli:	umbrella	parapluie		parapluie
passiv	'pas:iv	passive	passiv	passive	passif
potatis	po'tatis	potato		potato	patate
självständig	'ʃjɛ:lstændɪg	independent	selbstständig		
skriva	'skri:vɑ	to write	schreiben		
skärm	ʃæ:rm	screen	Schirm		
smart	smɑ:t	smart	smart	smart	
smink	smi:ŋk	make-up	Schminke		
tydlig	'ty:dli:g	clear	deutlich		
tänka	'tɛ:ŋkɑ	to think	denken	think	
ursprung	'ɔ:s,prɔ:ŋ	origin	Ursprung		
växla	'vɛ:ksla	to change	wechseln		
ägg	ɛ:g	egg	Ei	egg	œuf
äta	'ɛ:tɑ	to eat	essen	eat	
öst	ø:st	east	Ost(en)	east	est
överraska	ø:vɛ:'rɑ:skɑ	to surprise	überraschen		
börja	'bɔ:ʒjɑ	to begin	(profile word)		
flicka	'fli:kɑ	girl	(profile word)		
Sverige	'sʌ:ʒɛrjɛ	Sweden	(profile word)		
tråkig	'tro:kɪg	boring	(profile word)		

(Continued on next page)

Table A.2 – continued from previous page

Stimulus	Transcription	Translation	German	English	French
älska	'elska	to love	(profile word)		

## Appendix B

# Description and results of the Simon task

The Simon task was largely based on, but not identical to, the version used by Bialystok et al. (2004, Study 1). The participants were instructed to press the left button on the response pad (marked 'X') as fast as possible when a blue rectangle appeared on the screen and the right button (marked 'O') when a red rectangle appeared. The task consisted of a total of 28 trials, in fourteen of which the stimulus was presented on the same side as the required button response ('congruent'), whereas in the other fourteen trials, the side of the stimulus presentation did not match the response side ('incongruent'). Stimuli were presented in the same order for all participants. The Simon task was administered with E-Prime 2.0 (Schneider et al., 2002) and responses were recorded using Cedrus RB-834 response pads.

Each trial consisted of a fixation phase, during which a cross ('+', Courier New, 18 pt) was displayed in the centre of the screen for 800 ms, followed by a blank screen (250 ms). After the fixation phase, a blue or red rectangle appeared on the left or the right side of the screen and remained visible until the participant pushed a button. Response latencies were recorded from stimulus onset onwards. Intertrial intervals lasted 1,000 ms.

Before the actual Simon task, a training run with 8 stimuli took place. After this training run, participants could notify the experimenters in

case it emerged that they had not fully understood the instructions. Participants had to score perfectly on the training run before they could proceed to the actual Simon task. If they made a mistake, they needed to redo the training run until they attained a perfect score.

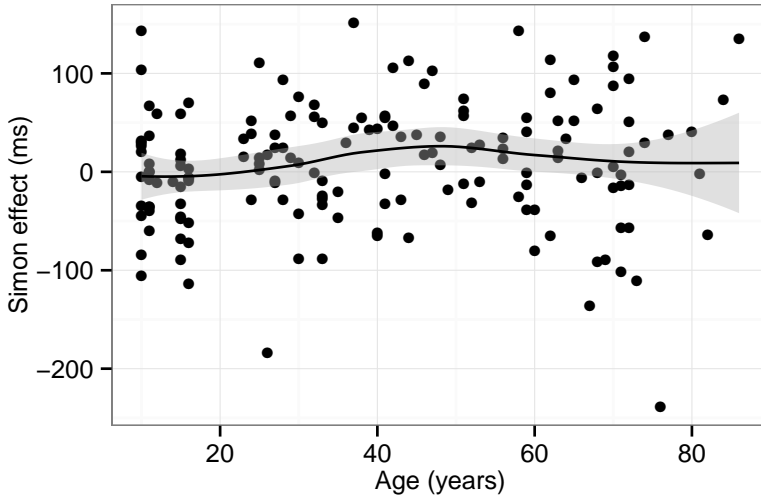
I computed the proportion of correct answers for each participant in both conditions, i.e. ‘congruent’ and ‘incongruent’. Additionally, I computed their median response latencies for the correct responses in both conditions separately. I chose to work with medians rather than with means as the former are substantially more robust to outliers. The Simon effect was expressed as the difference score (in milliseconds) between the median response latency in the incongruent condition and the median response latency in the congruent condition. Three participants had very low accuracy scores in both conditions. I assumed that, despite making it through the training run, these participants had simply applied the wrong response rule and therefore corrected their scores manually. One participant did not complete the Simon task, leaving a total of 162 participants with valid data.

The participants’ performance on the Simon task is summarised in Table B.1. Note that the Simon effect is negligibly small in terms of both accuracy and response latency. In fact, 74 of 162 participants have a faster median response latency in the incongruent condition than in the congruent condition. Figure B.1 additionally shows that the Simon effect in terms of latency does not show a clear age-related development; the accuracy scores were at ceiling and were not plotted.

A mixed-effect model in which the individual responses were modelled in function of a fixed effect of condition (congruent vs. incongruent), by-participant random intercepts and by-participants random slopes for the condition effect showed a significant but minute difference between both conditions in terms of accuracy (about one percentage point). However, mixed-effect models either revealed an inverse effect of condition on response latency with response latencies on incongruent trials being about 13 ms shorter than on congruent trials or no effect of condition whatsoever (less than a 1 ms difference between both conditions) depending on the outlier exclusion criteria. It can only be concluded that the Simon task failed to produce the canonical Simon effect and that its results hence cannot be used to represent the participants’ cognitive control ability.

**Table B.1:** Summary data for the Simon task ( $n = 162$ ).

		Range		Median	Mean	SD
		Lower	Upper			
Latency (ms)	Congruent	295	1024	474	491	125
	Incongruent	299	817	478	499	109
Difference score (ms)		-239	151	9	8	63
Accuracy (%)	Congruent	50	100	100	97	7
	Incongruent	57	100	100	96	6
Difference score (pp)		-36	21	0	1	7

**Figure B.1:** Lifespan development of the Simon effect.



# References

- Ackerman, P. L., 2000. Domain-specific knowledge as the “dark matter” of adult intelligence: Gf/Gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences* 55B: P69–P94.
- Ackerman, P. L., M. E. Beier and M. O. Boyle, 2002. Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General* 131: 567–589.
- Adams, A.-M. and S. E. Gathercole, 1995. Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research* 38: 403–414.
- Adams, A.-M. and S. E. Gathercole, 1996. Phonological working memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology* 49A: 216–233.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19: 716–723.
- Allen, D., 1992. *Oxford Placement Test*. Oxford: Oxford University Press, 2nd edn.
- Atkinson, R. C. and R. M. Shiffrin, 1968. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (eds), *The psychology of learning and motivation: Advances in research and theory*, vol. 2, 89–195. New York: Academic Press.
- Baayen, R. H., 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

- Baayen, R. H., D. J. Davidson and D. M. Bates, 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412.
- Baddeley, A. D., 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4: 417–423.
- Baddeley, A. D., 2002. Is working memory still working? *European Psychologist* 7: 85–97.
- Baddeley, A. D., 2003a. Working memory and language: An overview. *Journal of Communication Disorders* 36: 189–208.
- Baddeley, A. D., 2003b. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience* 4: 829–839.
- Baddeley, A. D., S. E. Gathercole and C. Papagno, 1998. The phonological loop as a language learning device. *Psychological Review* 105: 158–173.
- Baddeley, A. D. and G. J. Hitch, 1974. Working memory. In G. H. Bower (ed.), *The psychology of learning and motivation: Advances in theory and research*, vol. 8, 47–89. New York: Academic Press.
- Baddeley, A. D. and G. J. Hitch, 2000. Development of working memory: Should the Pascual-Leone and the Baddeley and Hitch models be merged? *Journal of Experimental Child Psychology* 77: 128–137.
- Baddeley, A. D., V. Lewis and G. Vallar, 1984. Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology* 36A: 233–252.
- Baddeley, A. D. and R. H. Logie, 1999. Working memory. The multiple-component model. In A. Miyake and P. Shah (eds), *Models of working memory: Mechanisms of active maintenance and executive control*, 28–61. New York: Cambridge University Press.
- Baltes, P. B., 1987. Theoretical propositions of life-span development psychology: On the dynamics between growth and decline. *Developmental Psychology* 23: 611–626.
- Baltes, P. B., U. M. Staudinger and U. Lindenberger, 1999. Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology* 50: 471–507.

- Bannert, R., 1981. Referat av diskussionen i sektionen Talperceptionsforskning och nordisk hörförståelse. In Elert (1981a), 37–45.
- Barr, D. J., R. Levy, C. Scheepers and H. J. Tily, 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–278.
- Bates, D., M. Maechler and B. Bolker, 2013. lme4: Linear mixed-effect models using S4 classes. R package, version 0.999999-2. URL <http://cran.r-project.org/package=lme4>
- Bayliss, D. M., C. Jarrold, A. D. Baddeley and D. M. Gunn, 2005. The relationship between short-term memory and working memory: Complex span made simple? *Memory* 13: 414–421.
- Beijering, K., C. Gooskens and W. Heeringa, 2008. Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm. *Linguistics in the Netherlands* 25: 13–24.
- Berent, I. and C. A. Perfetti, 1995. A rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review* 102: 146–184.
- Berger, T., 2003. Slovaks in Czechia—Czechs in Slovakia. *International Journal of the Sociology of Language* 2003: 19–39.
- Berndt, A., 2003. *Sprachenlernen im Alter. Eine empirische Studie zur Fremdspracheragogik*. Munich: Iudicium.
- Berthele, R., 2004. Vor lauter Linguisten die Sprache nicht mehr sehen—Diglossie und Ideologie in der deutschsprachigen Schweiz. In H. Christen (ed.), *Dialekt, Regiolekt und Standardsprache im sozialen und zeitlichen Raum: Beiträge zum 1. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen, Marburg/Lahn, 5.–8. März 2003*, 111–136. Vienna: Edition Praesens.
- Berthele, R., 2008. Dialekt-Standard Situationen als embryonale Mehrsprachigkeit. Erkenntnisse zum interlingualen Potenzial des Provinzlerdaseins. *Sociolinguistica* 22: 87–107.
- Berthele, R., 2011. On abduction in receptive multilingualism: Evidence from cognate guessing tasks. *Applied Linguistics Review* 2: 191–219.

- Berthele, R. and A. Lambelet, 2009. Approche empirique de l'intercompréhension : répertoires, processus et résultats. *Lidil* 39: 151–162.
- Berthele, R. and J. Vanhove, forthcoming. Entre jeunes barbes et vieux de la vieille. Usages du répertoire plurilingue dans une tâche d'intercompréhension à travers les âges. Accepted for publication in *Bulletin suisse de linguistique appliquée*.
- Besner, D., 1987. Phonology, lexical access in reading, and articulatory suppression: A critical review. *Quarterly Journal of Experimental Psychology* 39A: 467–478.
- Bialystok, E. and F. I. M. Craik (eds), 2006. *Lifespan cognition. Mechanisms of change*. New York: Oxford University Press.
- Bialystok, E., F. I. M. Craik, C. Grady, W. Chau, R. Ishii, A. Gunji and C. Pantev, 2005a. Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage* 24: 40–49.
- Bialystok, E., F. I. M. Craik, R. Klein and M. Viswanathan, 2004. Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging* 19: 290–303.
- Bialystok, E., F. I. M. Craik and G. Luk, 2008. Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34: 859–873.
- Bialystok, E. and G. Luk, 2012. Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition* 15: 397–401.
- Bialystok, E., M. M. Martin and M. Viswanathan, 2005b. Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism* 9: 103–119.
- Birdsong, D., 2006. Age and second language acquisition and processing: A selective overview. *Language Learning* 56: 9–49.
- Blake, J., W. Austin, M. Cannon, A. Lisus and A. Vaughan, 1994. The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development* 17: 91–107.

- Bø, I., 1978. *Ungdom og naboland: En undersøkelse av skolens og fjernsynets betydning for nabospråkforståelsen*. Stavanger: Rogalandsforskning.
- Brändle, M., 1986. Language teaching for the ‘young-old’. *Babel* 21: 17–21.
- Braunmüller, K. and L. Zeevaert, 2001. Semikommunikation, rezeptive Mehrsprachigkeit und verwandte Phänomene: Eine bibliographische Bestandsaufnahme. Arbeiten zur Mehrsprachigkeit/Working Papers in Multilingualism B-19, Sonderforschungsbereich 538 *Mehrsprachigkeit*, University of Hamburg, Hamburg.
- Braver, T. and D. Barch, 2002. A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience & Biobehavioral Reviews* 26: 809–817.
- Bravo, M. A., E. H. Hiebert and P. D. Pearson, 2007. Tapping the linguistic resources of Spanish/English bilinguals: The role of cognates in science. In R. K. Wagner, A. Muse and K. Tannenbaum (eds), *Vocabulary development and its implications for reading comprehension*, 140–156. New York: Guilford.
- Breiman, L., 1996. Bagging predictors. *Machine Learning* 24: 123–140.
- Breiman, L., 2001. Random forests. *Machine Learning* 45: 5–32.
- Broerse, A. C. and E. J. Zwaan, 1966. The information value of initial letters in the identification of words. *Journal of Verbal Learning and Verbal Behavior* 5: 441–446.
- Brysbaert, M., 2013. LexTALE\_FR: A fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica* 53: 23–37.
- Brysbaert, M., M. Buchmeier, M. Conrad, A. M. Jacobs, J. Bölte and A. Böhl, 2011. The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology* 58: 412–424.

- Brysbart, M. and B. New, 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41: 977–990.
- Bugg, J. M., E. L. DeLosh, D. B. Davalos and H. P. Davis, 2007. Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Aging, Neuropsychology, and Cognition* 14: 155–167.
- Campoy, G., 2008. The effect of word length in short-term memory: Is rehearsal necessary? *Quarterly Journal of Experimental Psychology* 61: 724–734.
- Carton, A. S., 1971. Inferencing: A process in using and learning language. In P. Pimsleur and T. Quinn (eds), *The psychology of second language learning: Papers from the Second International Congress of Applied Linguistics, Cambridge, 8–12 September 1969*, 45–58. London: Cambridge University Press.
- Cattell, R. B., 1963. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* 54: 1–22.
- Cattell, R. B., 1971. *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Cenoz, J., 2001. The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. In Cenoz et al. (2001), 8–20.
- Cenoz, J., B. Hufeisen and U. Jessner (eds), 2001. *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives*. Clevedon: Multilingual Matters.
- Cenoz, J., B. Hufeisen and U. Jessner (eds), 2003. *The multilingual lexicon*. Dordrecht: Kluwer.
- Chaoju, T. and V. J. Van Heuven, 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119: 709–732.
- Chertkow, H., V. Whitehead, N. Phillips, C. Wolfson, J. Atherton and H. Bergman, 2010. Multilingualism (but not always bilingualism)

- delays the onset of Alzheimer disease: Evidence from a bilingual community. *Alzheimer Disease & Associated Disorders* 24: 118–125.
- Cheung, H., 1996. Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology* 32: 867–873.
- Christen, H., M. Guntern, I. Hove and M. Petkova, 2010. *Hochdeutsch in aller Munde. Eine empirische Untersuchung zur gesprochenen Standardsprache in der Deutschschweiz*. Stuttgart: Steiner.
- Clark, H. H., 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12: 335–359.
- Cohen, J., 1983. The cost of dichotomization. *Applied Psychological Measurement* 7: 249–253.
- Cohen, J., 1994. The Earth is round ( $p < .05$ ). *American Psychologist* 49: 997–1003.
- Coleman, E. B., 1964. Generalizing to a language population. *Psychological Reports* 14: 219–226.
- Colom, R., I. Rebollo, A. Palacios, M. Juan-Espinosa and P. Kyllonen, 2004. Working memory is (almost) perfectly predicted by  $g$ . *Intelligence* 32: 227–296.
- Connine, C. M., D. G. Blasko and D. Titone, 1993. Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language* 32: 193–210.
- Connor, L. T., A. Spiro, III, L. K. Obler and M. L. Albert, 2004. Change in object naming ability during adulthood. *Journal of Gerontology: Psychological Sciences* 59B: P203–P209.
- Conway, A. R. A., N. Cowan, M. F. Bunting, D. J. Theriault and S. R. B. Minkoff, 2002. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence* 30: 163–183.
- Conway, A. R. A., M. J. Kane, M. F. Bunting, D. Z. Hambrick, O. Wilhelm and R. W. Engle, 2005. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12: 769–786.

- Conway, A. R. A., M. J. Kane and R. W. Engle, 2003. Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences* 7: 547–552.
- Costa, A., A. Caramazza and N. Sebastián-Gallés, 2000. The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1283–1296.
- Costa, A., M. Hernández, J. Costa-Faidella and N. Sebastián-Gallés, 2009. On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition* 113: 135–149.
- Cristia, A., A. Seidl, C. Vaughn, R. Schmale, A. Bradlow and C. Floccia, 2012. Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology* 3: 479.
- Cutler, A., 2005. Lexical stress. In D. B. Pisoni and R. E. Remez (eds), *The handbook of speech perception*, 264–289. Malden, MA: Blackwell.
- Cutler, A. and T. Otake, 2002. Rhythmic categories in spoken-word recognition. *Journal of Memory and Language* 46: 296–322.
- Cutler, A., N. Sebastián-Gallés, O. Soler-Vilageliu and B. Van Ooijen, 2000. Constraints of vowels and consonants on lexical selection: Cross-linguistic evidence. *Memory & Cognition* 28: 746–755.
- Cutler, A., B. Van Ooijen and D. Norris, 1999. Vowels, consonants and lexical activation. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville and A. C. Bailey (eds), *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, 1–7 August 1999*, vol. 3, 2053–2056. Berkeley, CA: Department of Linguistics, University of California.
- Daneman, M. and P. A. Carpenter, 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19: 450–466.
- Daneman, M. and R. Case, 1981. Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology* 17: 367–378.

- Daneman, M. and P. M. Merikle, 1996. Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review* 3: 422–433.
- De Angelis, G., 2007. *Third or additional language acquisition*. Clevedon: Multilingual Matters.
- De Groot, A. M. B. and R. Keijzer, 2000. What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning* 50: 1–56.
- Deary, I. J., 2001. Human intelligence differences: A recent history. *Trends in Cognitive Sciences* 5: 127–130.
- Deary, I. J., L. Penke and W. Johnson, 2010. The neuroscience of human intelligence differences. *Nature Reviews Neuroscience* 11: 201–211.
- Deary, I. J., F. M. Spinath and T. C. Bates, 2006. Genetics of intelligence. *European Journal of Human Genetics* 14: 690–700.
- Deary, I. J., S. Strand, P. Smith and C. Fernandes, 2007. Intelligence and educational achievement. *Intelligence* 35: 13–21.
- DeKeyser, R., 2012. Age effects in second language learning. In S. M. Gass and A. Mackey (eds), *The Routledge handbook of second language acquisition*, 442–460. London: Routledge.
- DeKeyser, R., I. Alfi-Shabtay and D. Ravid, 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31: 413–438.
- Delsing, L.-O. and K. Lundin Åkesson, 2005. *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska*. Copenhagen: Nordiska ministerrådet.
- Diependaele, K., K. Lemhöfer and M. Brysbaert, 2013. The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology* 66: 843–863.

- Dijkstra, T., J. Grainger and W. J. Van Heuven, 1999. Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language* 41: 496–518.
- Dijkstra, T. and W. J. B. Van Heuven, 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* 5: 175–197.
- Doetjes, G. and C. Gooskens, 2009. Skriftsprogets rolle i den dansk-svenske talesprogsforståelse. *Språk och stil* 19: 105–123.
- Dressler, C., M. S. Carlo, C. E. Snow, D. August and C. E. White, 2011. Spanish-speaking students' use of cognate knowledge to infer the meaning of English words. *Bilingualism: Language and Cognition* 14: 243–255.
- Duyck, W., D. Vanderelst, T. Desmet and R. J. Hartsuiker, 2008. The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review* 15: 850–855.
- Elert, C.-C. (ed.), 1981a. *Internordisk sprakförståelse. Föredrag och diskussioner vid ett symposium på Rungstegaard utanför Köpenhamn den 24–26 mars 1980*. Umeå: Umeå University.
- Elert, C.-C., 1981b. *Ljud och ord i svenskan 2*. Stockholm: Almqvist & Wiksell.
- Emmorey, K., G. Luk, J. E. Pyers and E. Bialystok, 2008. The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science* 19: 1201–1206.
- Ender, A., 2007. *Wortschatzerwerb und Strategieneinsatz bei mehrsprachigen Lernenden: Aktivierung von Wissen und erfolgreiche Verknüpfung beim Lesen auf Verständnis in einer Fremdsprache*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Engle, R. W., 2002. Working memory as executive attention. *Current Directions in Psychological Science* 11: 19–23.
- Engle, R. W., S. W. Tuholski, J. E. Laughlin and A. R. A. Conway, 1999. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General* 128: 309–331.

- Engstrand, O., 1990. Illustrations of the IPA: Swedish. *Journal of the International Phonetic Association* 20: 42–44.
- Eriksen, B. A. and C. W. Eriksen, 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* 16: 143–149.
- Faraway, J. J., 2006. *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Boca Raton, FL: Chapman & Hall/CRC.
- Flynn, J. R., 1987. Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin* 101: 171–191.
- Garlén, C. (ed.), 2003. *Svenska språknämndens uttalsordbok*. Stockholm: Svenska språknämnden & Norstedts.
- Gathercole, S. E., 1999. Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences* 3: 410–419.
- Gathercole, S. E. and A. D. Baddeley, 1993. *Working memory and language*. Hove: Erlbaum.
- Gathercole, S. E., S. J. Pickering, B. Ambridge and H. Wearing, 2004. The structure of working memory from 4 to 15 years of age. *Developmental Psychology* 40: 177–190.
- Gelman, A. and H. Stern, 2006. The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician* 60: 328–331.
- Gibson, M. and B. Hufeisen, 2003. Investigating the role of prior foreign language knowledge: Translating from an unknown into a known foreign language. In Cenoz et al. (2003), 87–102.
- Gigerenzer, G., 2004. Mindless statistics. *Journal of Socio-Economics* 33: 587–606.
- Gollan, T. H., R. I. Montoya, C. Cera and T. C. Sandoval, 2008. More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language* 58: 787–814.

- Gooskens, C., 2007a. Contact, attitude and phonetic distance as predictors of inter-Scandinavian communication. In J. Eloy and T. Ó hlfearnáin (eds), *Near languages - Collateral languages. Actes du colloque international réuni à Limerick, du 16 au 18 juin 2005*, 99–109. Paris: L'Harmattan.
- Gooskens, C., 2007b. The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development* 28: 445–467.
- Gooskens, C. and W. Heeringa, forthcoming. The role of dialect exposure in receptive multilingualism. Accepted for publication in *International Journal of Multilingualism*.
- Gooskens, C., W. Heeringa and K. Beijering, 2008. Phonetic and lexical predictors of intelligibility. *International Journal of Humanities and Arts Computing* 2: 63–81.
- Gooskens, C. and S. Kürschner, 2010. Hvilken indflydelse har danske stød og svenske ordaccenter på den dansk-svenske ordforståelse? In C. Falk, A. Nord and R. Palm (eds), *Svenskans beskrivning 30. Förhandlingar vid Trettionde sammankomsten för svenskans beskrivning. Stockholm den 10 och 11 oktober 2008*, 82–91. Stockholm: Department of Scandinavian Languages, Stockholm University.
- Gooskens, C., S. Kürschner and R. Van Bezooijen, 2011. Intelligibility of Standard German and Low German to speakers of Dutch. *Dialectologia* Special issue, II: 35–63.
- Gooskens, C. and R. Van Bezooijen, 2006. Mutual comprehensibility of written Afrikaans and Dutch: Symmetrical or asymmetrical? *Literary and Linguistic Computing* 21: 543–557.
- Gordon-Salant, S., 2005. Hearing loss and aging: New research findings and clinical implications. *Journal of Rehabilitation Research and Development* 42: 9–24.
- Gottfredson, L. S., 1997a. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence* 24: 13–23.

- Gottfredson, L. S., 1997b. Why *g* matters: The complexity of everyday life. *Intelligence* 24: 79–132.
- Gow, D. W., Jr., J. Melvold and S. Manuel, 1996. How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology and processing. In *Proceedings of the International Conference on Spoken Language Processing*, 66–69. Philadelphia.
- Graham, J. W., 2009. Missing data analysis: Making it work in the real world. *Annual Review of Psychology* 60: 549–576.
- Gray, J. R. and P. M. Thompson, 2004. Neurobiology of intelligence: Science and ethics. *Nature Reviews Neuroscience* 5: 471–482.
- Grüter, T. and M. Crago, 2012. Object clitics and their omission in child L2 French: The contributions of processing limitations and L1 transfer. *Bilingualism: Language and Cognition* 15: 531–549.
- Haastrup, K., 1991. *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tübingen: Gunter Narr.
- Hakuta, K., E. Bialystok and E. Wiley, 2003. Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science* 14: 31–38.
- Haugen, E., 1966. Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36: 280–297.
- Ház, É., 2005. *Deutsche und Niederländer. Untersuchungen zur Möglichkeit einer unmittelbaren Verständigung*. Hamburg: Dr. Kovač.
- Hebb, D., 1942. The effect of early and late brain injury upon test scores and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society* 85: 275–292.
- Heeringa, W., 2004. *Measuring dialect pronunciation differences using Levenshtein distance*. Ph.D. thesis, University of Groningen.
- Heitz, R. P., T. S. Redick, D. Z. Hambrick, M. J. Kane, A. R. A. Conway and R. W. Engle, 2006. Working memory, executive function, and general fluid intelligence are not the same. *Behavioral and Brain Sciences* 29: 135–136.

- Herdina, P. and U. Jessner, 2002. *A dynamic model of multilingualism: Perspectives of change in psycholinguistics*. Clevedon: Multilingual Matters.
- Hilchey, M. D. and R. M. Klein, 2011. Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review* 18: 625–658.
- Hilton, N. H., C. Gooskens and A. Schüppert, 2013a. The influence of non-native morphosyntax on the intelligibility of a closely related language. *Lingua* 137: 1–18.
- Hilton, N. H., C. Gooskens, A. Schüppert, R. Van Bezooijen and V. J. Van Heuven, 2013b. The influence of non-native morphosyntax and phonology on the intelligibility of a closely related language. In J.-M. Tirkkonen and E. Anttikoski (eds), *Proceedings of the 24th Scandinavian Conference of Linguistics*, 211–226. Joensuu: University of Eastern Finland.
- Hitch, G. J., 2006. Working memory in children: A cognitive approach. In Bialystok and Craik (2006), 112–127.
- Honaker, J., G. King and M. Blackwell, 2012. Amelia: Amelia II: A program for missing data. R package, version 1.6.4.  
URL <http://cran.r-project.org/package=Amelia>
- Horn, J. L., 1982. The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F. I. M. Craik and S. Trehub (eds), *Aging and cognitive processes*, 237–278. New York: Plenum.
- Horn, J. L. and R. B. Cattell, 1963. Age differences in fluid and crystallized intelligence. *Acta Psychologica* 26: 107–129.
- Horn, J. L. and R. B. Cattell, 1966. Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology* 57: 253–270.
- Hothorn, T., K. Hornik, C. Strobl and A. Zeileis, 2013. party: A laboratory for recursive partytioning. R package, version 1.0-6.  
URL <http://cran.r-project.org/package=party>

- Hufeisen, B. and N. Marx (eds), 2007. *EuroComGerm – Die sieben Siebe: Germanische Sprachen lesen lernen*. Aachen: Shaker.
- Impe, L., D. Geeraerts and D. Speelman, 2008. Mutual intelligibility of standard and regional Dutch language varieties. *International Journal of Humanities and Arts Computing* 2: 101–117.
- Jaeger, T. F., 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59: 434–446.
- Jaeger, T. F., P. Graff, W. Croft and D. Pontillo, 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15: 281–320.
- Jakimik, J., R. A. Cole and A. I. Rudnicky, 1985. Sound and spelling in spoken word recognition. *Journal of Memory and Language* 24: 165–178.
- James, A. R., 1983. Transferability and dialect phonology: Swabian–English. In A. R. James and B. Kettemann (eds), *Dialektphonologie und Fremdsprachenerwerb*, 173–197. Tübingen: Gunter Narr.
- Jarvis, S. and A. Pavlenko, 2008. *Crosslinguistic influence in language and cognition*. New York: Routledge.
- Jensen, J. B., 1989. On the mutual intelligibility of Spanish and Portuguese. *Hispania* 72: 848–852.
- Johnson, K., 2008. *Quantitative methods in linguistics*. Malden, MA: Blackwell.
- Johnson, R. L. and M. E. Eisler, 2012. The importance of the first and last letter in words during sentence reading. *Acta Psychologica* 141: 336–351.
- Johnson, W. and T. J. Bouchard, Jr., 2005. The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence* 33: 393–416.
- Johnson, W., T. J. Bouchard, Jr., R. F. Krueger, M. McGue and I. I. Gottesman, 2004. Just one *g*: Consistent results from three test batteries. *Intelligence* 32: 95–107.

- Johnson, W. and I. Gottesman, 2006. Clarifying process versus structure in human intelligence: Stop talking about fluid and crystallized. *Behavioral and Brain Sciences* 29: 136–137.
- Johnson, W., J. Te Nijenhuis and T. J. Bouchard, Jr., 2008. Still just 1 *g*: Consistent results from five tests batteries. *Intelligence* 36: 81–95.
- Jones, D. M., R. W. Hughes and W. J. Macken, 2006. Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language* 54: 265–281.
- Jones, D. M., R. W. Hughes and W. J. Macken, 2007. The phonological store abandoned. *Quarterly Journal of Experimental Psychology* 60: 505–511.
- Jones, D. M., W. J. Macken and A. P. Nicholls, 2004. The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 656–674.
- Jones, H. E. and H. S. Conrad, 1933. The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genetic Psychology Monographs* 13: 223–298.
- Just, M. A. and P. A. Carpenter, 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99: 122–149.
- Kane, M. J., M. K. Bleckley, A. R. A. Conway and R. W. Engle, 2001. A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General* 130: 169–183.
- Kane, M. J., L. H. Brown, J. C. McVay, P. J. Silvia, I. Myin-Germeys and T. R. Kwapil, 2007. For whom the mind wanders, and when. *Psychological Science* 18: 614–621.
- Kane, M. J., D. Z. Hambrick and A. R. A. Conway, 2005. Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin* 131: 66–71.

- Kellerman, E., 1977. Toward a characterisation of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin* 2: 58–145.
- Kellerman, E., 1983. Now you see it, now you don't. In S. Gass and L. Selinker (eds), *Language transfer in language learning*, 112–134. Rowley, MA: Newbury House.
- Kessels, R. P. C., D. Hobbel and A. Postma, 2007. Aging, context memory and binding: A comparison of “what, where and when” in young and older adults. *International Journal of Neuroscience* 117: 795–810.
- Kessler, B., 1995. Computational dialectology in Irish Gaelic. In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 60–66. San Francisco, CA: Morgan Kaufmann.
- Kievit, R. A., W. E. Frankenhuis, L. J. Waldorp and D. Borsboom, 2013. Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology* 4: 513.
- Klein, H. G. and T. D. Stegmann (eds), 2000. *EuroComRom – Die sieben Siebe: Romanische Sprachen sofort lesen können*. Aachen: Shaker.
- Kormos, J. and A. Sáfár, 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11: 261–271.
- Kray, J. and U. Lindenberger, 2007. Fluide Intelligenz. In J. Brandtstädter and U. Lindenberger (eds), *Entwicklungspsychologie der Lebensspanne*, 194–220. Stuttgart: W. Kohlhammer.
- Kuperman, V. and J. A. Van Dyke, 2013. Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance* 39: 802–823.
- Kürschner, S., C. Gooskens and R. Van Bezooijen, 2008. Linguistic determinants of the intelligibility of Swedish words among Danes. *International Journal of Humanities and Arts Computing* 2: 83–100.
- Kyllonen, P. and R. Christal, 1990. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 14: 389–433.

- Kyllonen, P. C., W. C. Tirre and R. E. Christal, 1991. Knowledge and processing speed as determinants of associative learning. *Journal of Experimental Psychology: General* 120: 57–79.
- Lee, H.-W., K. Rayner and A. Pollatsek, 2001. The relative contribution of consonants and vowels to word identification during reading. *Journal of Memory and Language* 44: 189–205.
- Lemhöfer, K. and M. Broersma, 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* 44: 325–343.
- Lemhöfer, K., T. Dijkstra, H. Schriefers, R. H. Baayen, J. Grainger and P. Zwitserlood, 2008. Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34: 12–31.
- Lemhöfer, K. and T. Dijkstra, 2004. Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition* 32: 533–550.
- Levenshtein, V. I., 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10: 707–710.
- Lotto, L. and A. M. B. De Groot, 1998. Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning* 48: 31–69.
- Lubliner, S. and E. H. Hiebert, 2011. An analysis of English-Spanish cognates as a source of general academic language. *Bilingual Research Journal* 34: 76–93.
- Luce, P. A. and D. B. Pisoni, 1998. Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing* 19: 1–36.
- Luk, G., E. de Sa and E. Bialystok, 2011. Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition* 14: 588–595.
- Lukatela, G. and M. T. Turvey, 2000. An evaluation of the two-cycles model of phonology assembly. *Journal of Memory and Language* 42: 183–207.

- Mangold, M., 2005. *Das Aussprachewörterbuch*. Mannheim: Dudenverlag, 6th edn.
- Marks, E. A., D. R. Moates, Z. S. Bond and L. Vázquez, 2002. Vowel mutability: The case of monolingual Spanish listeners and bilingual Spanish–English listeners. *Southwest Journal of Linguistics* 21: 73–99.
- Marslen-Wilson, W. D., 1987. Functional parallelism in spoken word-recognition. *Cognition* 25: 71–102.
- Marslen-Wilson, W. D. and A. Welsh, 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10: 29–63.
- Martin, K. I. and N. C. Ellis, 2012. The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition* 34: 379–413.
- Martin-Rhee, M. M. and E. Bialystok, 2008. The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition* 11: 81–93.
- Marx, N., 2007. Interlinguales Erschließen von Texten in einer unbekannt-germanischen Fremdsprache. *Zeitschrift für Fremdsprachenforschung* 18: 165–182.
- Marx, N., 2011. Reading across the Germanic languages: Is equal access just wishful thinking? *International Journal of Bilingualism* 16: 467–483.
- Maurud, O., 1976. *Nabospråksförståelse i Skandinavien: En undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige*. Stockholm: Nordiska rådet.
- McClelland, J. L. and J. L. Elman, 1986. The TRACE model of speech perception. *Cognitive Psychology* 18: 1–86.
- Meißner, F.-J., 1999. Das mentale Lexikon aus der Sicht der Mehrsprachigkeitsdidaktik. *Grenzgänge* 6: 62–80.
- Meißner, F.-J. and H. Burk, 2001. Hörverstehen in einer unbekannt-romanischen Fremdsprache und methodische Implikationen für den

- Tertiärsprachenerwerb. *Zeitschrift für Fremdsprachenforschung* 12: 63–110.
- Moates, D. R. and E. A. Marks, 2012. Vowel mutability in print in English and Spanish. *The Mental Lexicon* 7: 326–350.
- Möller, R., 2011. Wann sind Kognaten erkennbar? Ähnlichkeit und synchrone Transparenz von Kognatenbeziehungen in der germanischen Interkomprehension. *Linguistik online* 46: 79–101.
- Möller, R. and L. Zeevaert, 2010. “Da denke ich spontan an Tafel”: Zur Worterkennung in verwandten germanischen Sprachen. *Zeitschrift für Fremdsprachenforschung* 21: 217–248.
- Morton, J., 1969. Interaction of information in word recognition. *Psychological Review* 76: 165–178.
- Morton, J. B. and S. N. Harper, 2007. What did Simon say? Revisiting the bilingual advantage. *Developmental Science* 10: 719–726.
- Morton, J. B. and S. N. Harper, 2009. Bilinguals show an advantage in cognitive control—the question is why. *Developmental Science* 12: 502–503.
- Müller, M., L. Wertenschlag, S. Gerhartl, A. Halilbasic, I. Kaiser, E. Peyer, N. Shafer and R. Berthele, 2009. *Chunsch druus? Schweiz-erdeutsch verstehen – die Deutschschweiz verstehen*. Bern: Schulverlag.
- Müller-Lancé, J., 2003. A strategy model of multilingual learning. In Cenoz et al. (2003), 117–132.
- Muñoz, C. (ed.), 2006. *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
- Neisser, U., G. Boodoo, T. J. Bouchard, Jr., A. W. Boykin, N. Brody, S. J. Ceci, D. F. Halpern, J. C. Loehlin, R. Perloff, R. J. Sternberg and S. Urbina, 1996. Intelligence: Knowns and unknowns. *American Psychologist* 51: 77–101.
- New, B., M. Brysbaert, J. Veronis and C. Pallier, 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28: 661–677.

- Nieuwenhuis, S., B. U. Forstmann and E.-J. Wagenmakers, 2011. Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience* 14: 1105–1107.
- Nisbett, R. E., J. Aronson, C. Blair, W. Dickens, J. Flynn, D. F. Halpern and E. Turkheimer, 2012. Intelligence: New findings and theoretical developments. *American Psychologist* 67: 130–159.
- Nooteboom, S. G. and M. J. Van der Vlugt, 1988. A search for a word-beginning superiority effect. *Journal of the Acoustical Society of America* 84: 2018–2032.
- Norris, D., 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52: 189–234.
- Norris, D., 2006. The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review* 113: 327–357.
- Norris, D. and J. M. McQueen, 2008. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review* 115: 357–395.
- Oberauer, K., R. Schulze, O. Wilhelm and H.-M. Süß, 2005. Working memory and intelligence—Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin* 131: 61–65.
- Odlin, T., 1989. *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- Otwinowska-Kasztelanic, A., 2011. Awareness and affordances: Multilinguals versus bilinguals and their perception of cognates. In G. De Angelis and J.-M. Dewaele (eds), *New trends in crosslinguistic influence and multilingualism research*, 1–18. Bristol: Multilingual Matters.
- Paap, K. R. and Z. I. Greenberg, 2013. There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology* 66: 232–258.
- Pajak, B., 2010. Perceptual advantage from generalized linguistic knowledge. In S. Ohlsson and R. Catrambone (eds), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

- Papagno, C., T. Valentine and A. D. Baddeley, 1991. Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language* 30: 331–347.
- Papagno, C. and G. Vallar, 1992. Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *Quarterly Journal of Experimental Psychology* 44A: 47–67.
- Park, D. C. and D. Payer, 2006. Working memory across the adult lifespan. In Bialystok and Craik (2006), 128–142.
- Peereman, R., S. Dufour and J. S. Burt, 2009. Orthographic influences in spoken word recognition: The consistency effect in semantic and gender categorization tasks. *Psychonomic Bulletin & Review* 16: 363–368.
- Peirce, C. S., 1934. *Pragmatism and pragmaticism*. Cambridge, MA: Harvard University Press.
- Perre, L. and J. C. Ziegler, 2008. On-line activation of orthography in spoken word recognition. *Brain Research* 1188: 132–138.
- Perry, C. and J. C. Ziegler, 2002. On the nature of phonological assembly: Evidence from backward masking. *Language and Cognitive Processes* 17: 31–59.
- Portocarrero, J. S., R. G. Burrell and P. J. Donovick, 2007. Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology* 22: 415–422.
- R Core Team, 2013. R: A language and environment for statistical computing. Software, version 2.15.3.  
URL <http://www.r-project.org/>
- Rabbitt, P. and M. Anderson, 2006. The lacunae of loss? Aging and the differentiation of cognitive abilities. In Bialystok and Craik (2006), 331–343.
- Raven, J. C., 1962. *Advanced progressive matrices, set II*. London: H. K. Lewis.

- Reinfried, M., 1998. Transfer beim Erwerb einer weiteren romanischen Fremdsprache: Prinzipielle Relevanz und methodische Integration in den Fremdsprachenunterricht. In F.-J. Meißner and M. Reinfried (eds), *Mehrsprachigkeitsdidaktik: Konzepte, Analysen, Lehrerfahrungen mit romanischen Fremdsprachen*, 23–44. Tübingen: Gunter Narr.
- Repovš, G. and A. D. Baddeley, 2006. The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience* 139: 5–21.
- Rey-Debove, J. and A. Rey (eds), 1993. *Le Nouveau Petit Robert*. Paris: Dictionnaires Le Robert.
- Riggs, K. J., J. McTaggart, A. Simpson and R. P. J. Freeman, 2006. Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology* 95: 18–26.
- Ringbom, H., 1987. *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.
- Ringbom, H., 1992. On L1 transfer in L2 comprehension and L2 production. *Language Learning* 42: 85–112.
- Ringbom, H., 2007. *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters.
- Ringbom, H. and S. Jarvis, 2009. The importance of cross-linguistic similarity in foreign language learning. In M. H. Long and C. J. Doughty (eds), *Handbook of language teaching*, 106–118. Chichester: Wiley Blackwell.
- Rohde, T. E. and L. A. Thompson, 2007. Predicting academic achievement with cognitive ability. *Intelligence* 35: 83–92.
- Romani, C., S. McAlpine, A. Olson, E. Tsouknida and R. Martin, 2005. Length, lexicality, and articulatory suppression in immediate recall: Evidence against the articulatory loop. *Journal of Memory and Language* 52: 398–415.
- Rubin, J., 1975. What the “good language learner” can teach us. *TESOL Quarterly* 9: 41–51.

- Ruchkin, D. S., J. Grafman, K. Cameron and R. S. Berndt, 2003. Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences* 26: 709–728.
- Rudner, M. and J. Rönnerberg, 2008. The role of the episodic buffer in working memory for language processing. *Cognitive Processing* 9: 19–28.
- Salthouse, T. A., 1994. The aging of working memory. *Neuropsychology* 8: 535–543.
- Salthouse, T. A., 1996. The processing-speed theory of adult age differences in cognition. *Psychological Review* 103: 403–428.
- Salthouse, T. A., 2000. Aging and measures of processing speed. *Biological Psychology* 54: 35–54.
- Salthouse, T. A., 2006. Aging of thought. In Bialystok and Craik (2006), 274–284.
- Salthouse, T. A. and R. L. Babcock, 1991. Decomposing adult age differences in working memory. *Developmental Psychology* 27: 763–776.
- Salthouse, T. A. and E. J. Meinz, 1995. Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences* 50B: P297–P306.
- Scaltritti, M. and D. A. Balota, 2013. Are all letters really processed equally and in parallel? Further evidence of a robust first letter advantage. *Acta Psychologica* 144: 397–410.
- Schaie, K., 1996. *Intellectual development in adulthood. The Seattle Longitudinal Study*. Cambridge: Cambridge University Press.
- Scharloth, J., 2006. Schweizer Hochdeutsch – schlechtes Hochdeutsch? In C. Dürscheid and M. Businger (eds), *Schweizer Standarddeutsch: Beiträge zur Varietätenlinguistik*, 81–96. Tübingen: Gunter Narr.
- Schielzeth, H. and W. Forstmeier, 2009. Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology* 20: 416–420.

- Schiller, P. H., 1966. Developmental study of color-word interference. *Journal of Experimental Psychology* 72: 105–108.
- Schmidely, J., M. A. Ezquerro and C. H. González, 2001. *De una a cuatro lenguas. Intercomprensión Románica: del español al portugués, al italiano y al francés*. Madrid: Acro Libros.
- Schmidt, K.-H. and P. Metzler, 1992. *Wortschatztest (WST)*. Weinheim: Beltz Test.
- Schmitt, H., 2007. The case for the epsilon symbol ( $\epsilon$ ) in RP DRESS. *Journal of the International Phonetic Association* 37: 321–328.
- Schneider, W., A. Eschman and A. Zuccolotto, 2002. *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.
- Schüppert, A., 2011. *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish*. Ph.D. thesis, University of Groningen.
- Schüppert, A. and C. Gooskens, 2011. Investigating the role of language attitudes for perception abilities using reaction time. *Dialectologia* Special issue, II: 119–140.
- Schüppert, A., N. H. Hilton and C. Gooskens, forthcoming. Swedish is beautiful, Danish is ugly. Investigating the link between language attitudes and intelligibility. Accepted for publication in *Linguistics*.
- Schüppert, A., J. C. Ziegler, K. Magnusson, H. Juul, K. Holmqvist and C. Gooskens, 2010. Activation of L1 orthography in spoken word recognition of a closely related L2. Evidence from ERP. Poster presented at Experimental Approaches to Perception and Production of Language Variation, 11–12 November, Groningen.
- Service, E., 1992. Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology* 45A: 21–50.
- Service, E. and V. Kohonen, 1995. Is the relationship between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics* 16: 155–172.
- Shah, P. and A. Miyake, 1996. The separability of working memory resources for spatial thinking and language processing: An individual

- differences approach. *Journal of Experimental Psychology: General* 125: 4–27.
- Shiotsu, T. and C. J. Weir, 2007. The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing* 24: 99–128.
- Simmons, J. P., L. D. Nelson and U. Simonsohn, 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22: 1359–1366.
- Simon, J. R., 1969. Reactions toward the source of stimulation. *Journal of Experimental Psychology* 81: 174–176.
- Simon, J. R. and A. M. Small, Jr., 1969. Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology* 53: 433–435.
- Simpson, E. H., 1951. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 13: 238–241.
- Singer, T., P. Verhaeghen, P. Ghisletta, U. Lindenberger and P. B. Baltes, 2003. The fate of cognition in very old age: Six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging* 18: 318–331.
- Singleton, D. and D. Little, 1984. A first encounter with Dutch: Perceived language distance and language transfer as factors in comprehension. In L. Mac Mathúna and D. Singleton (eds), *Language across cultures: Proceedings of a symposium held at St Patrick's College, Dublin, 8–9 July 1983*, 259–270. Dublin: Irish Association for Applied Linguistics.
- Singleton, D. and L. Ryan, 2004. *Language acquisition: The age factor*. Clevedon: Multilingual Matters, 2nd edn.
- Smith, E. E. and J. Jonides, 1999. Storage and executive processes in the frontal lobes. *Science* 283: 1657–1661.
- Smith, M. K., 2005. *Development of visual-verbal integration in working memory during childhood and its relation to language and reading*. Ph.D. thesis, Georgia State University, Atlanta, GA.

- Spearman, C., 1927. *The abilities of man*. Oxford: Macmillan.
- Sternberg, R. J., 1997. The concept of intelligence and its role in lifelong learning and success. *American Psychologist* 52: 1030–1037.
- Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin and A. Zeileis, 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9: 307.
- Strobl, C., A.-L. Boulesteix, A. Zeileis and T. Hothorn, 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8: 25.
- Strobl, C., T. Hothorn and A. Zeileis, 2009a. Party on! A new, conditional variable-importance measure for random forests available in the party package. *R Journal* 1: 14–17.
- Strobl, C., J. Malley and G. Tuz, 2009b. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14: 323–348.
- Stroop, J. R., 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18: 643–662.
- Süß, H.-M., K. Oberauer, W. W. Wittmann, O. Wilhelm and R. Schulze, 2002. Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence* 30: 261–288.
- Swarte, F., A. Schüppert and C. Gooskens, 2013. Do speakers of Dutch use their knowledge of German while processing written Danish words? *Linguistics in the Netherlands* 30: 146–159.
- Tagliamonte, S. A. and R. H. Baayen, 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24: 135–178.
- Tainturier, M.-J., M. Tremblay and A.-R. Lecours, 1992. Educational level and the word frequency effect: A lexical decision investigation. *Brain and Language* 43: 460–474.

- Teleman, U., 1981. Om förståelse i allmänhet och om förståelse via grannspråken i synnerhet. Referat och kommentar till diskussionen i sektionen Den lexikala-semantiska komponenten i den nordiska hör-förståelsen. In Elert (1981a), 102–120.
- Tewes, U. (ed.), 1991. *HAWIE-R. Hamburg-Wechsler Intelligenztest für Erwachsene, Revision 1991: Handbuch und Testanweisungen*. Bern: Hans Huber.
- Turner, M. L. and R. W. Engle, 1989. Is working memory capacity task dependent? *Journal of Memory and Language* 28: 127–154.
- Unsworth, N., R. P. Heitz, J. C. Schrock and R. W. Engle, 2005. An automated version of the operation span task. *Behavior Research Methods* 37: 498–505.
- Uttl, B. and P. Graf, 1997. Color-word Stroop test performance across the adult life span. *Journal of Clinical and Experimental Neuropsychology* 19: 405–420.
- Van Bezooijen, R. and C. Gooskens, 2005a. How easy is it for speakers of Dutch to understand Frisian and Afrikaans, and why? *Linguistics in the Netherlands* 22: 13–24.
- Van Bezooijen, R. and C. Gooskens, 2005b. Intertalig tekstbegrip. De begripelijkheid van Friese en Afrikaanse teksten voor Nederlandse lezers. *Nederlandse Taalkunde* 10: 129–152.
- Van Bezooijen, R., C. Gooskens and S. Kürschner, 2012. Deens is makkelijker voor Friezen dan voor Nederlanders – feit of fabel? In P. Boersma, G. T. Jensma and R. Salverda (eds), *Philologia Frisica anno 2008. Lêzings fan it achttjinde Frysk Filologekongres fan de Fryske Akademy op 10, 11 en 12 desimber 2008*, 286–298. Leeuwarden: Afûk/Fryske Akademy.
- Van der Elst, W., M. P. Van Boxtel, G. J. P. Van Breukelen and J. Jolles, 2006. The Stroop Color-Word test. Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment* 13: 62–79.
- Van der Lubbe, R. H. J. and R. Verleger, 2002. Aging and the Simon task. *Psychophysiology* 39: 100–110.

- Van Heuven, V. J., 1985. Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America* 78: S21.
- Van Heuven, V. J., 2008. Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing* 2: 39–62.
- Van Ooijen, B., 1996. Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition* 24: 573–583.
- Van Wijnendaele, I. and M. Brysbaert, 2002. Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance* 28: 616–627.
- Vanhove, J., 2013. The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PLOS ONE* 8: e69172.
- Vanhove, J. and R. Berthele, forthcoming, a. Factoren bij het herkennen van cognaten in onbekende talen: algemeen of taalspecifiek? Accepted for publication in *Taal & Tongval*.
- Vanhove, J. and R. Berthele, forthcoming, b. Item-related determinants of cognate guessing in multilinguals. In G. De Angelis, U. Jessner and M. Kresić (eds), *Crosslinguistic influence and multilingualism*.
- Vanhove, J. and R. Berthele, forthcoming, c. The lifespan development of cognate guessing skills in an unknown related language. Accepted for publication in *International Review of Applied Linguistics in Language Teaching*.
- Verhaeghen, P., 2003. Aging and vocabulary scores: A meta-analysis. *Psychology and Aging* 18: 332–339.
- Verhaeghen, P. and T. A. Salthouse, 1997. Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin* 122: 231–249.

- Warter, P., 1995. Computersimulation von Wortverstehen am Beispiel mittelniederdeutsch-skandinavischer Sprachkontakte. In K. Braummüller (ed.), *Niederdeutsch und die skandinavischen Sprachen II*, 35–70. Heidelberg: Winter.
- Warter, P., 2001. Lexical identification and decoding strategies in inter-scandinavian communication. *Arbeiten zur Mehrsprachigkeit/Working Papers in Multilingualism B-21*, Sonderforschungsbereich 538 *Mehrsprachigkeit*, University of Hamburg, Hamburg.
- WASI, 1999. *WASI: Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.
- Welford, A. T., 1958. *Ageing and human skill*. London: Oxford University Press.
- Wells, J. C., 2000. *Longman Pronunciation Dictionary*. Harlow: Pearson Education, 2nd edn.
- West, R., 1999. Age differences in lapses of intention in the Stroop task. *Journal of Gerontology: Psychological Sciences* 54B: P34–P43.
- Williams, S. and B. Hammarberg, 1998. Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics* 19: 295–333.
- Wolff, H., 1959. Intelligibility and inter-ethnic attitudes. *Anthropological Linguistics* 1: 34–41.
- Wood, S. N., 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.
- Wood, S. N., 2013. mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package, version 1.7-24. URL <http://cran.r-project.org/package=mgcv>
- Ziegler, J. C., L. Ferrand and M. Montant, 2004. Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition* 32: 732–741.
- Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev and G. M. Smith, 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.