Vocabulary Testing and the Influence of Second Language on Third Language Vocabulary Acquisition

Thesis

How to cite:


For guidance on citations see FAQs.

© 2017 The Author

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.21954/ou.ro.0000ce71

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.

oro.open.ac.uk
Vocabulary Testing and the Influence of Second Language on Third Language Vocabulary Acquisition

Csaba - Zoltán Szabo
BA (Partium Christian University, RO)
MA (Swansea University, UK)

Thesis submitted to The Open University in fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisors:
Dr Lluisa Astruc
Dr Ursula Stickler
Dr Lina Adinolfi

2017
Declaration

I declare that this thesis presents my own work conducted during my scholarship as a full time doctoral research student at the Centre for Research in Education and Educational Technology (CREET) at The Open University, except where due acknowledgement is made, and that it has not been submitted previously to The Open University or to any other institution for a degree, diploma, or other qualification.

Certain parts of this thesis have been published or are being considered for publication:

- The development of the Romanian Word List and the Romanian Vocabulary Levels Test, presented in Chapter 3, was published in the proceedings of the 14th International Conference of the Department of Linguistics, University of Bucharest (Szabo, 2015)

- Study two, presented in Chapter 5, was published in the Eurasian Journal of Applied Linguistics (Szabo, 2016)


- A paper on academic achievement is also in preparation

The chapters of this thesis are more detailed versions of the published papers. The publications should be used for page references and citations.
Abstract

Learning and speaking foreign languages are regarded today as the norm. Simultaneously, empirical findings in the area of multilingualism have recently demonstrated that the mental lexicon of a language learner consists of intertwined systems where languages interact with each other, rather than remain separate entities. However, studies that explore how different lexicons are integrated and influence each other in the case of multilinguals are relatively scarce.

This thesis seeks to contribute to this gap in research reporting two empirical studies drawing on a vocabulary testing perspective. The studies investigate the impact of prior lexical knowledge on additional language learning of Hungarian native speakers, who speak Romanian as an L2 and English as an L3.

From a theoretical standpoint, the studies contest the traditional assumption that foreign language vocabulary acquisition can be explained solely by measures of frequency of word occurrence. Instead is contended that cross-linguistic similarities, namely cognates, provide a special bridge between languages and also need to be taken into account. At a more substantive level, it aims to explore (1) the relationship between learners’ Romanian and English written lexical knowledge; (2) the facilitatory effect of cognates; and (3) the implications of this for vocabulary assessments. To address these aims, Romanian versions of Nation’s Vocabulary Levels Test (VLT) and Vocabulary Size Test (VST) have been developed, assessed for validity, and applied.

The findings indicate that there is a strong connection between multilinguals’ Romanian L2 and English L3 lexica. Learners’ lexical proficiency can be described as a function of frequency, but is also influenced by cognateness. Therefore, word frequency and cognateness in conjunction can potentially increase test accuracy and validity, and enable a more in-depth understanding of vocabulary size and lexical accessibility. These conclusions are drawn from a vocabulary testing perspective, but further pedagogical implications, future directions for research and limitations are also offered.
Acknowledgements

First and foremost, I would like to express my gratitude to the Centre for Research in Education and Educational Technology (CREET) at The Open University and my supervisors for selecting me for this doctoral scholarship, without which this dream would have hardly materialised. The invaluable financial, academic, and personal support throughout these three and a half years have amalgamated into a truly enriching and rewarding experience. I have learnt a lot not just about vocabulary, but about myself as well.

Unquestionably, I am most grateful to my supervisors, Dr Lluisa Astruc, Dr Ursula Stickler, and Dr Lina Adinolfi for their encouragement, support, and dedication throughout. Giving me the freedom of embracing something of interest meant the world to me. Each and every one of you brought a wealth of distinct expertise, challenged me in different ways, and importantly, helped me to stay grounded and focused. I will remain indebted.

Special thanks to my former professors Dr Bökös Borbála, Dr Borbély Julianna, and Drd Gáll Laura at Partium Christian University for the motivating words and providing access to their students. Without such support, this research would have been impossible.

I am also thankful to the academic community, colleagues, and friends whom I met along this journey. First, Swansea University where my interest in lexical matters was ignited. Professors Tess Fitzpatrick and Paul Meara for their availability and approachability, constructive feedback, and practical, academic, and personal support whenever needed. Having the opportunity to attend and present at various conferences and events on vocabulary (Lexical Studies group at Cardiff/Swansea University; BAAL Vocab SIG) and multilingualism helped enormously in the shaping and sharpening of this thesis. Special thanks to fellow CREET students, you were of great support and I am glad we shared this experience together.

I would also like to thank the Santander Universities Network for the Mobility Grant, which enabled me to spend three months at the University of Barcelona and expand my research. I am indebted to Dr Imma Miralpeix, Dr Ana Pellicer-Sánchez, and Professor Carmen Muñoz for welcoming me to Barcelona, making me feel at home, and providing feedback on my second publication.

Writing a PhD can be a lonely and challenging experience. At darker times, when nothing made sense and I felt lost, both intellectually and emotionally, I had my friends, here in the UK and back home, to pull me back, cheer me up, and encourage and convince me time and time again that I am capable of doing it. There are not enough words to express my appreciation for being there for me and the many memories that we share. I love you all. I am especially grateful to Adam for living this with me and supporting me throughout.

Finally, my heartfelt thanks, and perhaps apology, to my mum and sister for my absence these years. I hope one day I can make up for it.
Contents

Abstract ................................................................................................................................. v
Acknowledgements ................................................................................................................ vii
Contents ............................................................................................................................... ix
List of Figures ...................................................................................................................... xii
List of Tables ......................................................................................................................... xiii
List of Abbreviations ............................................................................................................. xv

Chapter 1: Introduction ................................................................................. 1
  1.1 The prevalence of a multilingual dimension to lexical studies ......................... 2
  1.2 The prevalence of lexical research to language acquisition ............................... 8
  1.3 The aims and objectives of the thesis ................................................................. 10
  1.4 The structure and content of the thesis .............................................................. 15

Chapter 2: Literature Review ...................................................................... 17
  2.1 Vocabulary knowledge ...................................................................................... 17
  2.2 Words, dimensions of lexical knowledge, and the mental lexicon (ML) .......... 22
     2.2.1 The definition and characteristics of words .................................................. 22
     2.2.2 Dimensions of lexical knowledge ............................................................... 26
     2.2.3 The mental lexicon (ML) ............................................................................ 31
  2.3 Cross-linguistic similarities: cognates in vocabulary research ......................... 40
  2.4 Vocabulary knowledge and the influence of background languages ............... 48
  2.5 The effect of explicit instruction (EI) on cognates ............................................ 61
  2.6 Lexical knowledge and speed of lexical access ................................................. 72
  2.7 Third Language Acquisition: proficiency in multiple languages ...................... 75

Chapter 3: Measuring Vocabulary Size in the L2 ..................................... 80
  3.1 Cultural and educational context ..................................................................... 81
  3.2 Teaching and learning vocabulary in the Romanian context .............................. 85
  3.3 English frequency lists and measuring vocabulary size ................................... 87
  3.4 The development of the Romanian Word List (RWL) ...................................... 93
  3.5 The Vocabulary Levels Test (VLT) ................................................................ 103
  3.6 The Romanian Vocabulary Levels Test (RomVLT) .......................................... 107
  3.7 The Vocabulary Size Test (VST) and developing its L1 and L2 versions ........... 110
  3.8 Issues of reliability and validity ....................................................................... 115

Chapter 4: Study One (S1) ......................................................................... 120
  4.1 Objectives and Research Questions ................................................................. 120
  4.2 Design of Study One ....................................................................................... 124
7.2.2 Limitations relating to the dimensions of vocabulary knowledge ............... 276
7.2.3 Limitations relating to the tests employed..................................................... 277
7.2.4 Limitations relating to item characteristics.................................................... 278
7.2.5 Limitations relating to research design......................................................... 281
7.3 Final conclusions.............................................................................................. 284

References ............................................................................................................ 287
Appendices ......................................................................................................... 306
List of Figures

Figure 3.1 Ethnic minorities in Transylvania (2011 census) .......................................................... 82
Figure 3.2 RWL frequency distribution .................................................................................................. 99
Figure 4.1 RomVLT mean results on different levels ......................................................................... 137
Figure 4.2 VLT mean results on different levels .................................................................................. 137
Figure 4.3 Barplot for the Study and Control groups on the two tests .............................................. 146
Figure 4.4 Scatterplot with the VLT and RomVLT results ..................................................................... 150
Figure 4.5 Interaction between proficiency and cognateness on the VLT ............................................ 158
Figure 4.6 Interaction between proficiency and cognateness on the RomVLT ................................. 161
Figure 4.7 Interaction between proficiency and cognateness based on frequency on the VLT................. 162
Figure 4.8 Interaction between proficiency and cognateness based on frequency on the RomVLT .......... 162
Figure 5.1 ENVST mean scores by frequency ...................................................................................... 181
Figure 5.2 ROVST mean scores by frequency ...................................................................................... 182
Figure 5.3 Mean results on the ENVST and ROVST by year ................................................................. 187
Figure 5.4 Scatterplot with overall scores on the two tests ................................................................. 192
Figure 5.5 ENVST Item facility boxplot for the different levels by cognate status .............................. 196
Figure 5.6 ROVST Item facility boxplot for the different levels by cognate status .............................. 196
Figure 5.7 Item Facility distribution by language and cognate status .................................................. 198
Figure 5.8 ENVST mean cognate score interaction by proficiency ...................................................... 203
Figure 5.9 ROVST mean cognate score interaction by proficiency ...................................................... 205
Figure 5.10 Mean cognate score interaction by proficiency ................................................................. 206
Figure 5.11 Interaction between proficiency and cognateness based on frequency on the ENVST ......... 209
Figure 5.12 Interaction between proficiency and cognateness based on frequency on the ROVST ........ 209
Figure 5.13 Mean RT divided by language and cognate status .............................................................. 217
Figure 5.14 Interaction between proficiency and cognateness on the ENVST ...................................... 218
Figure 5.15 Interaction between proficiency and cognateness on the ROVST ...................................... 219
Figure 5.16 Interaction between proficiency and cognateness on the HUVST ...................................... 220
Figure 6.1 Distribution of words on the BNC/COCA frequency list by Romanian cognate status .......... 249
List of Tables

Table 3.1 The distribution of tokens and types in sections ................................................. 96
Table 3.2 Frequency comparison between the BNC and RWL ........................................ 99
Table 3.3 Comparison between the two versions of the RWL ........................................ 100
Table 3.4 The coverage of the RWL in the Literature section (2) .................................... 101
Table 3.5 Distribution of cognates on the tests .............................................................. 108
Table 4.1 Descriptive statistics for the VLT and RomVLT .............................................. 136
Table 4.2 Model of VLT scores by levels .......................................................... 140
Table 4.3 Model of RomVLT scores by levels ............................................................. 141
Table 4.4 Descriptive statistics of the two tests by year of study ................................... 143
Table 4.5 Descriptive statistics for the Study and Control groups on the two tests........ 147
Table 4.6 Cognate results on different frequency levels on the VLT ......................... 147
Table 4.7 Cognate results on different frequency levels on the RomVLT ....................... 148
Table 4.8 Descriptive statistics for cognate and non-cognate items ............................... 152
Table 4.9 Cognate results on the VLT by proficiency ................................................. 156
Table 4.10 Cognate results on different frequency levels on the RomVLT ................. 159
Table 4.11 Summary of answers to Questions 2 and 3 .............................................. 165
Table 4.12 Summary of answers to Questions 4 and 5 .............................................. 165
Table 5.1 Descriptive statistics for the ENVST and ROVST scores based on levels .... 183
Table 5.2 Model of ENVST scores by level .............................................................. 184
Table 5.3 Model of ROVST scores by level ............................................................... 185
Table 5.4 Descriptive statistics for the overall scores ............................................... 186
Table 5.5 Descriptive statistics for the ENVST based on groups and levels ............... 191
Table 5.6 Descriptive statistics for the ROVST based on groups and levels ............... 191
Table 5.7 Spearman’s correlation figures for the different levels ............................... 193
Table 5.8 Descriptive statistics for cognate and non-cognate items ......................... 195
Table 5.9 ENVST cognate results on different frequency levels ................................ 202
Table 5.10 ROVST cognate results on different frequency levels ................................ 204
Table 5.11 Model of ENVST scores by proficiency and cognateness ....................... 207
Table 5.12 Model of ROVST scores by proficiency and cognateness .......................... 208
Table 5.13 Descriptive statistics for the test scores and RT divided by proficiency .... 213
Table 5.14 Mean RT distributed by frequency ......................................................... 214
Table 5.15 Mixed model of RT on the ENVST and ROVST .................................... 216
Table 5.16 Mean RT based on the different item properties ..................................... 222
Table 5.17 Mean item facility based on the different test properties ....................... 223
Table 5.18 Descriptive statistics for self-rating questions (N = 54) ............................ 227
List of Abbreviations

AWL – Academic Word List
CG – Control Group
CLI – Cross-linguistic Influences
EI – Explicit Instruction
FL – Foreign Language
GPA – Grade Point Average
HUVST – Hungarian Vocabulary Size Test
L1 – First Language (mother-tongue or native-language)
L2 – Second Language (or Foreign Language, depending on context)
L3 – Third or Additional Language in the case of bi- or multilinguals
LA – Language Acquisition
LL – Language Learning
ML – Mental Lexicon
NS – Native Speaker
NNS – Non-Native Speaker
POS – Part of Speech
RQ – Research Question
RomVLT – Romanian Vocabulary Levels Test
ROVST – Romanian Vocabulary Size Test
RQ – Research Question
RWL – Romanian Word List
S1 – Study One in this thesis
S2 – Study Two in this thesis
SG – Study (Experimental) Group
TL – Target Language or Additional Language (the language under study)
UWL – University Word List
VLT – Vocabulary Levels Test
VS – Vocabulary Size
VST – Vocabulary Size Test
Chapter 1: Introduction

The past two decades have brought a significant upturn in the number of research studies that place vocabulary in the central foci of attention, thus making lexical studies a prosperous field of Applied Linguistics both in terms of the large variety of theoretical questions addressed and those involved, from researchers to educational practitioners. Besides scholars who belong to the applied or cognitive and sociolinguistic cohorts, theoretical or formal linguists also conjoin in emphasising the importance of lexical aspects of language. Chomsky’s Minimalist Programme (1995), for example, contends that the human language faculty, which comprises innate language universals previously conceptualised as structures defined by Universal Grammar, in actuality, consists of different lexica across languages and ‘all that language acquisition involves is the learning of the lexicon’ (Mitchell and Myles, 2004: 66). From an historical perspective, Milton and Donzelli also state that ‘arguments are beginning to emerge at a theoretical level that vocabulary knowledge is the driver in the acquisition of a foreign language’ (2013: 441).

The primary reason for this is that vocabulary knowledge and development is a complex and multi-faceted area of cognition that is fundamental to understand and be understood. Therefore, uncovering the intricacies of the lexicon is a prerequisite for a detailed understanding of language acquisition (LA), especially in the case of second language comprehension, processing, and use, and in order to expedite language learning (LL). Despite the combinatorial potential of previous findings and the current issues of relevance these explore in relation to lexical knowledge, acquisition, and testing, from a learner-related perspective, the majority of empirical investigations can be described as monostratal. That is, vocabulary research is mainly concerned with the actual state or rate of uptake of vocabulary, how this can be measured in the target language (TL) and in what ways this is interconnected with acquisition in general. While a limited number of enquiries explore the role of the first language in second language acquisition (SLA), little is known
Chapter 1: Introduction

about the intricate nature of multilingualism from a lexical point of view.

Consequently, the major topical theme of this thesis is whether, and if so to what extent and how, the knowledge of multiple languages influences lexical knowledge in a TL, and by extension, the acquisition of this, and the prevalence of such factors to vocabulary testing. Although from a philosophical standpoint this doctoral research employs a positivist approach within a cognitivist theoretical framework, as the forthcoming chapters will demonstrate, I also adopt a conciliatory view under which formal linguistic (i.e. the proportion and distribution of inter-linguistic lexical similarities) and sociolinguistic (i.e. learners’ predisposition towards their background languages and the socio-cultural context where the learning takes place) properties of language form a dovetail. The following introductory section lays out the topicality of a multilingual perspective to lexical studies. This is ensued by the distinct aims and objectives of this thesis. Finally, the structure and content of the forthcoming chapters are presented.

1.1 The prevalence of a multilingual dimension to lexical studies

The ability to master and use several languages is seen today rather as the norm in most communities and societies globally. Together with this, the desire to learn foreign languages (FL) has enormously increased as well. This is instantiated by a US study, which confirms that 80% of Americans consider that children should be able to speak a second language (L2) fluently prior to entering university (Rivers, Robinson, Harwood and Brecht, 2013). Simultaneously, 84% of Europeans regard that everyone should be able to use a FL (European Commission Special Barometer [ECSB], 2006). Although some estimations reconcile this desire by asserting that most language users globally speak at least two languages (e.g. Aronin and Singleton, 2012; Crystal, 2003; Szubko-Sitarek, 2015), in actuality, only 56% of adults in the European Union report fluency in an additional language despite the widespread popularity of FL teaching (ECSB, 2006). What seems to be the main reason that works against additional LA is the seemingly insurmountable lexical knowledge
base that is necessary to function in a language. In Milton and Donzelli’s view ‘communicability and comprehension are likely to be compromised if the learner’s vocabulary knowledge is deficient’ (2013: 441). Therefore, establishing different ways to ameliorate the unbalanced standards of language teaching and bolster our understanding of LA appears to be a ubiquitous conundrum.

Since many languages share linguistic similarities, it is somewhat surprising, however, that until recently researchers have tried to address important questions of how human languages work, how we acquire, teach, and assess them effectively from a monolingual perspective, thus disregarding the similarities and interaction between the known languages in the case of bilinguals or even multilinguals. Recently, the realisation that using multiple languages is a natural phenomenon in many parts of the world (e.g. Aronin and Singleton, 2012; Herdina and Jessner, 2002) has led to a growing interest in bilingualism initially, and multilingualism later on. Still, most current research on first and second (or additional) LA disregards the interaction between such languages and other potential background languages. This is historically due to prejudice towards multilinguals on the part of the monoglossic elite and researchers opting for the ‘pure’ object of study (Chomsky, 1986), mostly for reasons of convenience (Szubko-Sitarek, 2015).

Compared to learning Latin, Greek, or French, which in the past was a reflection of social status (in Europe) and speakers of these languages were characterised as part of a social elite, today, being bi- or even multilingual is defined by historical events, changes in our needs, technologisation, increasing short and long term mobility, immigration, and language policies. This is particularly so in the European context, where inter alia, the post-war period brought about several alterations in different countries’ borders, causing heritage languages to face a decline and forcing a number of communities to adapt to a new culture and learn a different language. Some examples of these are the Austro-Hungarian territories affected by The Treaty of Trianon in 1920 with Transylvania being the largest territory lost
Chapter 1: Introduction

to Romania, Welsh and Gaelic as heritage languages within the United Kingdom, or Catalan and Basque in Spain.

Nowadays, learning different languages and settling down in different cultures is more a reflection of the younger generation’s needs and lifestyle, and thus, forms a fundamental part of a common European identity (Otwinowska, 2016). Multilingualism, as such, cannot be regarded as a homogeneous characteristic owing to the fact that the constellation of the known languages, users’ proficiency in and predisposition towards those languages will vary from one individual to another.

This variation is due to the historical, socio-economic, and political changes in recent years. On the other hand, it is connected to issues relating to globalisation and the position of English as an International Language or Lingua Franca. Since English is the primary language in several countries, one of the official languages in over 70 countries and territories, its current status as a global language is unsurprising (Crystal, 2003).

Teaching, learning, and communicating in English, however, are not analogous across borders. One clear distinction is when it is acquired as an L1 or L2 in an English-speaking country and, of course, learnt in less natural settings, where it is seen simply as a FL. This has led to English becoming part of most people’s Dominant Language Constellations (DLC) (e.g. Aronin and Singleton, 2012), in which it is either a global language used in television or online, language of popular culture, trade, and education, or one of several official languages (e.g. as in India or the European Union), a locally important, or a state language. According to Singleton and Aronin (2007), although English is often of major importance, it has increasingly become just one part of the language repertoire, but as such, it also forms an inseparable part of the individuals’ DLC. By definition, DLC constitutes of an individual’s languages that are to some extent frequently, or dominantly, used as opposed to language repertoire, which apart from these dominant languages, also includes other languages that are not frequently used, but the individual has at least some knowledge of
these (Aronin and Singleton, 2012).

Taking into account the fact that the DLC differs from multilingual to multilingual and that several languages share a number of linguistic similarities, it is theorised that the process of LA and lexical knowledge of multilinguals will vary considerably based on their knowledge and experience of learning previous languages (e.g. Cenoz, 2003; De Angelis, 2007). Furthermore, De Groot and Kejzer (2000), for example, state that the experience of learning a FL may aid the acquisition of subsequent languages by facilitating vocabulary development due to efficient learning strategies, cumulative learning experience, and prior linguistic knowledge. This premise is primarily based on the fact that English, albeit categorised as a Germanic language, 75% of its lexicon derives from other languages (Winford, 2003), such as Latin (17%) and French (45%) (Van Gelderen, 2006). This means that the 10,000 most common English words are largely of Romance origin, creating a strong lexical connection between English and other Latinate languages, such as Italian, French, Romanian, and Spanish (Otwinowska, 2016).

These compelling interconnections, cross-linguistic similarities as it were (e.g. Cenoz, 2003; De Angelis, 2007; Otwinowska, 2016), between certain languages led to the development and spread of Multilingual Language Acquisition (also termed Third Language Acquisition, TLA) as a separate branch of Applied Linguistics, with a growing number of dedicated conferences, such as the Multilingual Language Acquisition, Processing and Use, journals such as the International Journal of Multilingualism, and influential book editions (e.g. De Angelis, 2007; Szubko-Sitarek, 2015).

Compared to SLA, Multilingual Language Acquisition stems from the notion that languages are not stored as separate, individual entities, but language competence is being seen as a holistic concept of an intertwined system of multiple languages (e.g. Daryai-Hansen et. al, 2014; Falk and Bardel, 2010; Herdina and Jessener, 2002). In this view, ‘a multilingual is neither the sum of three or more monolinguals, nor a bilingual with an
additional language’ (De Angelis and Selinker, 2001: 44). Therefore, multilinguals are rather seen as individuals capable of acquiring and using several languages at varying degrees of proficiency. The known languages, then, are characterised by co-existence and mutual interconnections determined by multiple dimensions, such as self-identification with these languages, their origin, level of engagement at the individual level and so on (e.g. Jessner, 2008; Otwinowska, 2016).

This major shift in paradigm is also reflected in the language and education policies promoted by the Council of Europe (COE; 2007, 2010) in which plurilingualism occupies a focal position and seminal papers that advocate a Focus on Multilingualism in not just research practices, but pedagogical approaches as well (e.g. Cenoz, 2013; Gorter and Cenoz, 2016). These emphasise that inter-linguistic approaches and metalinguistic awareness should be encouraged in classrooms, as L2 learners have access to a plurilingual or pluricultural competence that involves all the languages available to them and this should be developed accordingly (COE, 2010).

One example of such approach is the FREPA (Framework of Reference for Pluralistic Approaches to Languages and Cultures) project, reviewed by Daryai-Hansen et al. (2014), which aims to complement the Common European Framework of Reference (CEFR). It describes how different macro-level initiatives integrate plurilingualism in their educational curricula and how distinct aspects of language knowledge, skills, and attitude can all be encouraged under the same aegis. They nevertheless emphasise a discrepancy between the implementation of this holistic concept and current teaching and research practices, and call for new assessment tools and research perspectives that espouse plurilinguistics.

It is important to define and differentiate between the terms multilingualism and plurilingualism. Multilingualism is the traditional Anglophone umbrella term used to define and describe both the individual and societal ability of using several languages, and the area of research that investigates this phenomena (e.g. De Angelis, 2007; Jessner, 2008). The
COE distinguishes between multilingualism, relating to cultures, and plurilingualism relating to individuals. As such, multilingualism denotes and encompasses different varieties of languages pertaining to one geographical area. The notion of plurilingualism (including pluricultural and plurilingual) is a neologism mostly used in European documents to refer to the individual capable of functioning in several languages and derives from French (Mackiewicz, 2002). The CEFR provides the following exhaustive definition of plurilingualism:

> the ability to use languages for the purposes of communication and to take part in intercultural interaction, where a person, viewed as a social agent, has proficiency of varying degrees, in several languages, and experience of several cultures. This is not seen as a superposition or juxtaposition of distinct competences, but rather as the existence of a complex or even composite competence on which the user may draw (COE, 2001: 168).

In accordance with the research in this area (e.g. De Angelis, 2007; Jessner, 2008; Otwinowska, 2016), throughout this thesis, the terms multilingual and multilingualism will be used interchangeably with plurilingual and plurilingualism to denote language users knowing more than two languages to varying degrees, and multilingualism per se will also refer to the broad area of research that investigates this. In conjunction with this, L2 will denote a second (or foreign) language that either chronologically in the acquisition or based on proficiency precedes the third language (L3), and Ln will refer to cases of multiple languages where the exact number, actual order of acquisition, or proficiency in that language is not of particular importance. Consistent with previous research will be the distinction that whereas the TL of learners or the language under study in SLA is the L2, in TLA, and adopted in this thesis, this is the L3. Thus, the L1 (mother tongue) and L2 will be specific background languages, and the L3 will be the language under investigation and learners’ TL, explicitly specified or irrespective of proficiency or order of acquisition.

Despite these recent conceptualisations, initiatives and, as Chapter 2 highlights, research findings in several areas mostly beyond LA, the most evident component of linguistic similarity that a multilingual’s repertoire will share, have been scarcely addressed
thus far. According to De Angelis (2007) such similarity is mainly and most easily observable on the lexical level. One reason for this is that English shares a large number of words with other languages, and thus, it follows that lexical congruencies across languages per se can be regarded as an underlying construct that influences LA and these effects will be pertinent to other areas as well, such as vocabulary processing, use, and testing.

1.2 The prevalence of lexical research to language acquisition

Vocabulary studies is a well-established branch of applied linguistics that chiefly focuses on the lexical knowledge and development of monolinguals and L2 learners alike in a principled and systematic way (e.g. Milton, 2013). As a result, numerous methods and assessment tools have been developed to facilitate the understanding of word knowledge, quantification and measurement of the lexicon ad hoc or over time, in the form of lexical uptake or attainment, diversity or richness, attrition or fossilisation. Such receptive and productive vocabulary size (VS) tests with a diverse range of formats, today are popular not just in academic circles, but also in educational settings (e.g. Meara and Miralpeix, 2016; Milton, 2009; Nation, 2013). The reasons behind this spring from a wide variety of some empirically substantiated and academically grounded arguments on the one hand, and some conventions that build on previous traditions without necessarily challenging these principles, on the other.

Amongst the commonly accepted postulations are those that regard vocabulary knowledge as a key factor in LL (e.g. Meara, 1980; Milton, 2009; Milton and Fitzpatrick, 2014; Nation, 2001, 2013; Read, 2000; Schmitt, 2010). Wilkins (1972) and Ellis (1997), inter alia, contend that lexis is indispensable for the acquisition of grammar. Similarly, Laufer and Sim (1985) posit that learners’ greatest barrier to L2 comprehension is lexical deficit, and this is more prevalent than grammatical, syntactic, or subject matter knowledge. Research into the cognitive processing in reading also conceptualises lexical skills at the heart of careful and expeditious reading (e.g. Bax, 2013; Khalifa and Weir, 2009). The
The prevalence of lexical research to language acquisition

operationalisation of this using eye-tracking indicates that ‘successful candidates [in a reading test] also showed better abilities at the lexical level’, leading Bax to conclude that ‘it is valuable for learners to work on their lexical knowledge’ (2013: 461). While Brown also considers that vocabulary acquisition is crucial to not just reading, but L2 acquisition in general as ‘lexical items are basic to all of the four skills’ (2007: 435), it has to be noted that LA is not monolithic but consists of various salient constituents that are per se dynamic, and interact synchronically and diachronically with each other on linguistic, cognitive, and social levels.

There is evidence that the four different language skills, as conceptualised today by the CEFR and operationalised in language assessments such as the International English Language Testing System (IELTS), namely the productive skills of writing and speaking (including interactive speaking in the CEFR) and the receptive skills of reading and listening can be largely determined as a function of vocabulary (e.g. Daller and Phelan, 2013; Milton, 2009; 2013). That is, the more words learners know, the better they will tend to be at these skills, and by extension, also in academia, regardless of whether the L1 or Ln is investigated (e.g. Daller and Wang, 2016; Daller and Phelan, 2013; Milton and Treffers-Daller, 2013). Moreover, there is also compelling evidence for learners needing to master a minimal number of words, so-called thresholds, in order to be able to communicate, read, and, overall, function in a FL (e.g. Milton, 2009; Milton and Donzelli, 2013; Milton and Fitzpatrick, 2014; Nation, 2001, 2013; Read, 2000; Schmitt, 2010), a task that has always represented a challenge for language learners and teachers.

By contrast, there are certain gaps that are yet to be subjected to empirical justifications, for example, the rate at which words are acquired, which words are more easily acquired than others, and what underlies this process; the extent to which lexical items are known, where this knowledge might derive from and why. There are several conventions that lie behind these elusive concepts, however, given the current advancements in research,
it seems timely to reconsider such traditionalised assumptions.

1.3 The aims and objectives of the thesis

The main, seldom-contested, aspect of vocabulary acquisition and testing is the issue of frequency of word occurrence. Languages are learnt incrementally and in this process, commonly used (i.e. both frequency of encounter and frequency of practice) words are likely to be acquired faster than infrequent ones (e.g. Milton, 2009; Nation, 2001; Read, 2000; Schmitt, 2010). The significance of this is best illustrated by Milton’s claim: ‘the importance of frequency in vocabulary learning is as near to a fact as it is possible to get in L2 acquisition’ (2009: 242), insofar as, lexical tests are largely built on the ‘laws’ of frequency (see Chapter 3). Although departing from this and similar traditionalised assumptions is not completely viable, this thesis initiates to challenge and problematise some of these conventions from at least two angles.

The first one is concerned with the issue of frequency of words in a language and whether this is the sole or most influential factor that contributes to LA or enables understanding the underlying issues that confound and mask lexical assessments and their interpretation. Since languages are interconnected on several levels and a certain proportion of lexis is shared across languages, it is proposed that in order to meaningfully encapsulate the concept of vocabulary knowledge and the intricacies of language testing, research needs to look beyond the TL and its properties.

There is clear benefit in exploring, identifying and, then, defining and modelling how prior lexical knowledge influences the development of vocabulary knowledge. Hence, this aspect of the thesis is concerned with lexical items that are shared between English and Romance languages. Such words are not necessarily mastered in accordance with their frequency, or other word characteristics (e.g. word length, part of speech [POS]) in English. Therefore, knowing, using, teaching, and even testing such words must be qualitatively and, maybe, quantitatively different in the case of multilinguals. Exploring and exposing these
differences could potentially contribute to the calibration of lexical tests and result in an enhanced, more subtle, and informed interpretation of the test scores and, ultimately, estimations of VS.

The second facet, which is strongly connected to the first, considers lexical knowledge and VS as a whole. The lexicon of the learner, as the ensuing will demonstrate on several levels, is a complex and dynamic entity. This complexity is further influenced by the interplay between an individual’s known languages. Consequently, a more profound comprehension of how words are stored and retrieved by multilinguals is considered to shed light in a fine level of granularity on the relationship between vocabulary size(s) in multiple languages.

This idea is founded in plurilingualism, the implementation of a multilingual focus in the classroom, and closing the discrepancy between research and practice. Should these concepts contribute to LA, the need for new tools and an understanding of how multiple VSs interact with one another becomes hardly questionable. Furthermore, it would seem unreasonable to believe that if performance on a lexical test can indicate language performance in the different language modalities, or academic achievement, then vocabulary knowledge in one language cannot be a good indicator of lexical knowledge in another FL. If vocabulary acquisition and natural languages are driven by the same mechanism or laws, i.e. frequency, then due to the same cognitive system and shared conceptual knowledge, VSs in additional languages should also share common characteristics and, perhaps, research in this direction can reveal and eventually model such linguistic and cognitive synergies.

In light of these composite issues and given the number of possible language and proficiency combinations, it also becomes evident that capturing the multidimensional nature of multilingual vocabulary knowledge would certainly be unachievable even in a PhD thesis. Subsequently, it seems logical to focus on one language constellation precisely, and then return to the wider context to illuminate this darker corner of SLA.
This thesis, therefore, sets out to investigate the above issues in relation to one specific cohort of multilinguals, who, by and large, share the same DLC. In particular, central to the research will be Hungarian native speakers (L1) living and studying in Transylvania, a macro-economic region of Romania, where the state language is Romanian and English is the most popular taught language (the cultural and educational context is further discussed in Chapter 3). As the aforementioned highlighted, English shares a large number of lexis with Latinate origin, hence, a substantive aim will be to inquire into the relationship between English as an L3 and a Romance L2. On a theoretical level, then, the aim will be to initiate a discussion and empirically explore influential factors of TLA, especially with regard to English and an already known Romance language. The ratiocination behind this is largely based on previous research suggesting that learners’ interlanguage will be influenced by the language that is typologically closer in a lexical sense to the TL (De Angelis, 2007; Singleton, 2003) which, in turn, can make vocabulary testing problematic.

Many scholars (e.g. Nation, 1983; Schmitt, Schmitt and Clapham, 2001; Milton, 2009; Laufer and Ravenhorst-Kalovski, 2010) posited that the vocabulary knowledge of speakers of languages originating from Latin may be overestimated due to the high number of similar words shared between these languages and English. However, no previous studies have systematically addressed the nature and extent of this overestimation in a multilingual context (i.e. beyond the influence of the mother-tongue). Since VS tests based on frequency levels are key means for evaluating learners’ lexical knowledge, the aim of the two empirical studies presented in this thesis is to explore how lexical similarities between background languages can be accounted for in English lexical assessments and how this is influenced by Romanian as an L2. A further aim, which is unique to this thesis, is to inquire into how multiple VSs are incorporated into the mental lexicon, which, consequently, adds to the current knowledge of the relationship between and the interaction of a Romance language and English lexica.
In order to investigate these aspects of the multilingual mental lexicon (ML), I address the following overarching research questions (RQ): (1) what is the relationship between Romanian L2 and English L3 lexical knowledge? (2) are there any differences in how interlinguistically congruent and incongruent items are recognised by learners in Romanian and English?

The two RQs necessarily lead to a broader inquiry that has been widely substantiated in other areas of research, but still represents a paucity in vocabulary studies: should speakers of more than two languages be considered as learners in their own right compared to monolingual L2 learners, and if this is in the affirmative, what would be the consequences for lexical testing? If prior lexical knowledge can represent a multisource and multistratal influence for LA and test development, it is entirely tenable that estimations of VS are much more complicated than previously surmised. The two empirical studies conducted consider addressing these questions from several perspectives.

First, to be able to compare Romanian and English vocabulary knowledge, compatible tools are required that produce reliable and valid results. Such tools should also be capable of accounting for cross-linguistic congruencies on the lexical level between the two languages. From a pedagogical and testing perspective, however, accounting for such similarities provides an incomplete picture. Therefore, I am also interested in whether English-Romanian congruent items (or cognates, see Chapter 2) are automatically recognised, or such similarities between the two languages should be introduced to learners via explicit instruction (EI) (see Milton and Donzelli, 2013 for a distinction on implicit and explicit learning). The benefit of this is that if prior lexical knowledge governs vocabulary recognition and learners capitalise on this knowledge automatically, then lexical acquisition is not solely frequency driven, which bears implications for testing. Conversely, if cognate recognition is susceptible of improvement by raising awareness of cross-lexical similarities, then the implementation of such classroom methods can enhance vocabulary development.
and inform pedagogical implications.

Second, since Romanian and English are surmised as learners’ interlanguages (e.g. De Angelis, 2007), i.e. not yet fully developed language systems, there is clear value in exploring the link between these languages. I will employ two different approaches in attempt to achieve this. From a within-subject perspective, I will explore whether there is a connection between lexical knowledge in the L2 and L3 overall and, specifically, at the level of cognates. Furthermore, if cognates are shared between the L2 and L3, it logically follows that cognate items are more easily recognised than non-cognates. This will be addressed in an item analysis, which, in turn, can facilitate a more in-depth understanding and nuanced interpretation of test scores.

Third, I also acknowledge that lexical proficiency can vary in these interlanguages and, therefore, it is entirely conceivable that learners might have to reach a certain proficiency level to benefit from the facilitatory effect of cognates. To investigate this, I explore different avenues that can potentially reveal more about the interplay between proficiency and cognate recognition.

Finally, the impetus of this thesis is the desire to develop and probe a framework that places the multilingual learner with varying language abilities, perceptions of their languages, and different LL experiences central to research. Therefore, in conjunction with specific language-related issues, a self-developed questionnaire will guide the understanding of salient learner-related factors, such as learners’ self-rated proficiency and confidence in their interlanguages, their perception of the link between their languages, and the age of onset of learning each of these languages. These, then, will form the locus of attention from a validity perspective. In particular, it will enable gathering more information about the language learners and drawing stronger conclusions on the relationship between lexical knowledge and learner-centric aspects, thus informing this exploration of the size of the multilingual lexicon.
In order to further bolster the validity of this framework and the importance of accounting for learners’ background languages in lexical testing, an external measure that carries meaning for both educators and learners is also explored. Namely, learners’ academic achievement will be compared to their lexical knowledge in the two languages under study and explored whether severally or in conjunction these can predict academic grades as a criterion-validity metric. The logical entailment of this is that drawing far-reaching conclusions on the employability of lexical tests is impossible without scrutinising what these purport to measure from a validity point of view (see Chapter 3).

1.4 The structure and content of the thesis

With these theoretical and practical goals in mind, Chapter 2 presents a systematic review and critical synthesis of the above-introduced concepts and expands these, reviews the relevant theories and findings within the areas of SLA and TLA, and more specifically, vocabulary studies, with a special focus on cross-linguistic similarities on the level of lexis. Analysing and interpreting the relationship between two languages from a VS perspective requires, in this case, comparable Romanian tools, for which a Romanian frequency list is indispensable.

Therefore, Chapter 3 provides a brief overview of how teaching Romanian vocabulary is conceptualised in different publications. For this to be complete, though, the cultural and educational background of the focus-group of this thesis is also afforded. Subsequently, the chapter fills a long-standing gap in the context of Romanian teaching and vocabulary testing, and proposes a novel frequency list that allows for the development of the Romanian versions of the lexical tests employed herein. Thus, Chapter 3 concludes with the deliberation and critical evaluation of the English tests and the compilation of their Romanian versions, relying on the Romanian Frequency List as a source. This deliberately and conveniently also encapsulates the instruments employed in the studies featured in this PhD research into one place for critical evaluation and scrutiny.
Chapter 1: Introduction

The two studies, in Chapter 4 and 5 respectively, are presented as a self-standing report each providing a detailed description of the formulated RQs, ethical consideration prior to the design of the studies, methodological aspects and decisions taken, and finally, presents the analysis and interpretation of the results. Since the two studies were conceptualised and operationalised incrementally, study two builds up on, further develops, and interprets the findings in study one, albeit largely exploring the same underlying aspects at a more fine-grained level.

The primary reason for this, in actuality, is the lack of research that compares receptive vocabulary knowledge in two typologically related languages in a systematic and principled fashion. Accordingly, in order to ameliorate for the paucity of empirical research in the area, it seemed justified to first explore the hypotheses in Chapter 4, and then, revisit these findings. This involved employing a different, and based on the nature of test-items, a meaning-based multiple-choice lexical recognition test, not just in the L2 and L3, but the L1 as well. Employing an L1 vocabulary test was required for the a priori consideration that, should the second study replicate and consequently also strengthen the findings in study one, introducing a temporal measure would open up the possibility of exploring participants’ speed of lexical access in conjunction with VS in the dominant languages. Subsequently, this would enable extrapolating beyond lexical scores and examine the relationship between lexicons on another, more fine-tuned, level.

Finally, Chapters 6 and 7 bring the results together for a holistic discussion and return to the wider perspective, exploring the contributions and implications of the present research, and how it relates to the broader context, while also acknowledging its limitations, and offering potential research directions in a similar vein.
Chapter 2: Literature Review

This chapter provides a broad overview of the most relevant research on vocabulary knowledge with a special focus on theories and empirical studies that have a multilingual approach. As the thesis investigates multilingual lexical knowledge from three different aspects, part one will point to key concepts in vocabulary and review studies that explored how vocabulary knowledge and size is influenced by prior lexical knowledge and language similarities, including cognates, and the reason why they deserve special attention. The second part takes a pedagogical perspective and reviews studies that have tried to address the facilitative effect of cognates and whether giving special attention to them enhances vocabulary uptake or LA in general. The last sections present and evaluate vocabulary studies employing a reaction time (RT) paradigm, and elaborate on the role of proficiency in multilinguality.

2.1 Vocabulary knowledge

Vocabulary knowledge, as a source of linguistic information that determines all other aspects of language processing and use, is complex and dynamic in its nature. This complexity is characterised by the various facets that play an important role in defining it (e.g. Nation, 2001; Milton, 2009, Schmitt, 2010). These facets include a wide range of characteristics, from establishing what a word or vocabulary is to the different aspects and degrees of lexical knowledge. Chapter 2 has introduced the concept of frequency as one of the influential factors in understanding vocabulary knowledge. However, other characteristics may include (e.g. Meara, 1992, 2010; Milton, 2009; Milton and Fitzpatrick, 2014; Nation, 2013; Schmitt, 2009):

- The definitions of words: single or multiword units; contextual or grammatical links to other words (collocations, phraseological units etc.); different types of units of count (types, tokens etc.); orthographic and phonological representation etc.

- Different dimensions of vocabulary knowledge: receptive and productive
The number of words known and amount of information that is known about these words: breadth and depth

The storage, organisation, and accessibility of known words: the mental lexicon (ML)

Speed, automaticity, and fluency: lexical access and processing

Each of these features have a fundamental impact in synchronicity on not just the understanding of vocabulary acquisition, but on the compilation and meaningful interpretation of lexical tests that try to tap into these different notions. It is not surprising, then, that pinning down the intricacies of the complexities of vocabulary knowledge into a coherent description or definition has perplexed researchers to date (Schmitt, 2014). As Milton and Fitzpatrick places it in an historical context:

Two thousand years later we are still wrestling with words and most recently we have been puzzling over the nature of word knowledge, and how and where this knowledge is stored. Knowing a word is an elusive concept and we are still unable to capture, in a simple description, everything that knowing a word might involve. Word knowledge, it seems, is complicated and it is hard to capture all of its many facets in a simple yet comprehensive definition (2014: 1).

There are several reasons for trying to understand and define vocabulary knowledge, however. For example, the ubiquity and significance of vocabulary in LA has led to a growing number of studies exploring the relationship between lexical knowledge and language competence, especially in English (e.g. Laufer and Goldstein, 2004; Milton, 2009, 2013). Although this is premised in that language ability necessarily comprises various components (Bachmann and Palmer, 1996), a large cohort of researchers now agree that vocabulary is a core element of language performance and processing, and therefore, ‘lexical knowledge is conceptually the most important component of variability in language ability’ (Treffers-Daller, 2011: 149). In a review of several studies, Milton (2013) extrapolates that there is a strong link between VS and performance in a FL, including all the four key language skills.

This sizable relationship between lexical dexterity and language performance has been addressed in the case of native speakers (NS) by Milton and Treffers-Daller (2013). They
contend that ‘[s]tudents with larger vocabularies tend to score higher in their assignments and exams and to obtain higher degree classifications than those with smaller vocabularies’ (p. 166). A clear link between L2 learners’ vocabulary knowledge and academic achievement has also been demonstrated (e.g. Daller and Wang, 2016; Daller and Phelan, 2013; Jarvis and Daller, 2013). Others indicated that there is also a relationship between vocabulary knowledge (both receptive and productive) and general intelligence (e.g. Anderson and Freebody, 1981; Daller and Phelan, 2013) or grammatical ability (e.g. Zimmerman, 2004). Milton and Alexiou (2009) compared receptive vocabulary scores against the progress on the CEFR levels as an indicator of overall language proficiency, and arrived to the conclusion that this is closely related to lexical dexterity. In particular, Spanish and Greek L2 learners’ English and French VSs explained around 70% of the variance of the CEFR levels.

Consequently, understanding the many aspects of vocabulary knowledge, determining factors of VS and its measurement is far from easy. However, previous research has gone some way in linking VS to other aspects of language knowledge. This also allows, inter alia, creating and interpreting admission and placement tests (e.g. Laufer and Nation, 1999; Meara and Jones, 1988). It enables capturing the rate of vocabulary acquisition, which subsequently helps educators and researchers to better inform curricular decisions and improve teaching methodologies (e.g. Milton, 2009; Nation, 2013). It could indicate learners, and other educational stakeholders, of the level they need to attain to be able to study at a foreign university or work abroad (e.g. Daller and Wang, 2016; Milton and Treffers-Daller, 2013). Moreover, an in-depth apprehension of lexical acquisition and vocabulary knowledge could contribute to improving LL software and other developments. In light of these possibilities, measuring VS has become the central foci in the past decades not just for vocabulary researchers, but teachers as well, leading to the development of a number of lexical tests, albeit mainly receptive.
Based on the number of studies and validation data, the most notable examples of such receptive tests (for a detailed description, see Chapter 3) are the Vocabulary Levels Test (VLT, Nation, 1983; Schmitt, Schmitt and Clapham, 2001), the Vocabulary Size Test (VST, Nation, 2006; Nation & Beglar, 2007), and the so-called yes/no tests, such as Meara and Milton’s X-Lex (2003) with their various incarnations. Productive tests would include adaptations of the above, such as the Productive Vocabulary Levels Test (PVLT, Laufer and Nation, 1999) novel tools, such as Lex30 (Meara and Fitzpatrick, 2000) or the Lexical Frequency Profile (LFP, Laufer and Nation, 2005), and a hybrid version of productive and receptive tests, such as the Computer Adaptive Test of Size and Strength (CATSS, Laufer & Goldstein, 2004). These ostensibly uncomplicated tests have already been employed in SLA to explore learners’ VS, to allow for monitoring of progress and allocating students to different proficiency groups in the case of placement tests.

The problem with these tests, however, is that although they measure receptive or productive lexical knowledge to some extent, the results cannot necessarily be compared across cultures, learning contexts, and individual language backgrounds, which, in turn, make it difficult to standardise them. As Cobb rightly states, ‘vocabulary tests can focus on either language or the learner’ (2000: 300) and in most research to date, the focus has been mainly on the language, disregarding other factors that can influence lexical knowledge. Recently, it has been argued, however, that due to the effect of existing lexical resources on additional LL and learners’ cumulative learning experience, the conceptualisation and operationalisation of a multifaceted lexical knowledge posits several challenges. Maybe this is the reason why measuring VS in multiple languages and exploring the relationship and interconnection between them is still an understudied area (e.g. De Angelis, 2007; Szubko-Sitarek, 2015).

There is growing evidence that the difference between monolinguals, bilinguals and multilinguals can be fundamental. For instance, Bialystok, Craik, and Luk (2008) found that
bilinguals from a variety of language backgrounds regularly outperform their monolingual counterparts on letter fluency (i.e. recalling as many words [e.g. animals] starting with a specified letter in a given language as possible) and word naming tasks (i.e. producing words based on a certain set of stimulus) due to their working memory, or executive control. Barac and Bialystok (2011) tested different bilingual groups (Chinese-English, French-English, and Spanish-English) and a monolingual English group on verbal and non-verbal tasks. Their conclusion indicates that on the executive control tasks (colour-shape task switching) the bilinguals have outperformed the monolinguals, regardless of language background. However, on the receptive and productive vocabulary tasks, only the Spanish-English bilinguals showed an advantage due to cross-linguistic influences (CLI).

Employing a different research paradigm to explore proficiency in L2 English, Molnar (2008) found that Hungarian-Romanian bilinguals outperformed the Hungarian monolingual group on the VLT, however, the Romanian monolingual group obtained significantly higher scores than both other groups due to their high proficiency in the typologically closer language, that is, Romanian. This supports Treffers-Daller’s (2011) claim about the importance of obtaining detailed information of participants’ lexicon, and as this thesis attempts to demonstrate, since background languages have a fundamental impact on the ML of multilingual learners, prior to establishing the relationship between VS and other competencies, an understanding of multiple VSs and the interaction between these is required.

As the second part of this chapter aims to highlight, the reason for this is that lexical similarities between certain languages have the potential to facilitate LL and thus, it can be argued that these should be taken into account to arrive at a clear and detailed picture of the ML.

The central similarity that is of concern herein is the issue of cognates, their potential influence on vocabulary acquisition and lexical tests in the case of multilinguals. Prior to
establishing the effect of cognates on language testing and why these special lexical items are regarded as a bridge between languages, it is necessary to explore some key issues in vocabulary research: the conceptualisation of the mental storage and the operationalisation of words and lexical knowledge in language testing thus far.

2.2 Words, dimensions of lexical knowledge, and the mental lexicon (ML)

2.2.1 The definition and characteristics of words

A key issue in lexical research is establishing the number of words a learner already knows or has to know according to their needs, their organisation in the ML, their relation to LA, and ultimately, identifying ways to explore this via tests that are both valid and reliable. Three underlying factors make this line of enquiry problematic. The first is the concept of words, or lexical items, and their definition and characteristics. Different research approaches define lexis from different perspectives. One example is the psycholinguistic approach, where words have to be strictly controlled in order to account for the different aspects of language knowledge, processing, and use. The best way to illustrate the wide range of these characteristics is through the MRC Psycholinguistic Database (Coltheart, 1981), which employs over 30 criteria to aid the selection of stimulus words. Some of these are:

- Number of letters, phonemes, syllables
- Familiarity, meaningfulness, concreteness and imageability ratings
- Morphemic and contextual status
- Syntactic category
- Frequency
- POS

Although the database provides a considerable number of word characteristics, the list is far from comprehensive or exhaustive as it features only individual lexical items. Corpus linguistics, on the other hand, allows researchers to investigate, for instance, how different
words ‘behave’ in different contexts, what their most frequent collocations are and if there exist any particular patterns for these, whether the words’ meaning and usage differs in spoken and written modes and so on. These examples highlight two important distinctions between how vocabulary is conceptualised, as a collection of words the learner knows, that is, as single lexical items or multiword units. According to Schmitt (2010) both of these conceptualisations have grounds in research, depending on the aspect of language or lexical knowledge that is being researched. The reason for this distinction lies in the fact that words in general are not always stored as single entries, but a considerable amount of language has been shown to be used (or retrieved), and therefore stored, holistically as multiword units, or as more frequently termed, formulaic sequences (e.g. Conklin and Schmitt, 2012; Schmitt, Grandage and Adolphs, 2004; Wray, 2000; 2002).

Another way to categorise lexis is as counting units in a discourse or for language tests. The problem language researchers, teachers, and learners face most often is that many aspects of language do not come in ‘conveniently sized packages that can be counted’ (Milton, 2009: 6). The qualities of an essay or a conversation, for example, make it difficult to assess them objectively. The assessor perhaps might not agree with the arguments or register selected by the student to convey his/her ideas. In a conversation or oral examination, external factors, such as too much noise or uncomfortable temperatures, and internal factors, such as an anxious student speaking very quietly, might influence the scoring, even if teachers do their best to remain objective and learners try to adhere to the marking criteria. Similarly, this lack of objectivity and the complexity of language knowledge makes it difficult to measure progress. In most schools today students are marked on scales, 1-10 in Romania (or 1-5 in Hungary, or A-F in the UK), for example, where 5 is a pass. Typically, advanced learners in a Romanian classroom receive around 8-10 on average throughout their secondary school. This, however, does not mean that students made no progress during these years or that a score of 5 for two individuals in the same classroom, or in different schools,
accounts to similar knowledge or language competencies. The CEFR attempts to rectify this by setting clear indicators for learners and teachers in relation to what needs to be achieved to progress from one level to the other. However, in the case of B1, for instance, it is still difficult to compare individual students not just in different countries in Europe, but even on national levels based on different hours of exposure, different language backgrounds, access to the target culture and so forth.

This is of relevance, for measures of VS allow to meaningfully quantify, and therefore, at least attempt to estimate lexical knowledge at different stages of development, compare individual and group knowledge, and even assess progress. The question is, then, how can words be quantified as units of language or text? In vocabulary research, a distinction is normally made between tokens, types, word families, lemmas (lemmata) and hapax legomena (entails words that appear only once in a corpus). Tokens or running words refer to all the words in a sentence, book or corpus, whereas, types refer to all the different lexis, counted only once. Tokens are mostly utilised in quantifying the size of a dictionary or a corpus, measuring the length of a university assignment, or counting the rate of speaking or reading. From a teaching or research perspective however, it is much more useful to understand how many different type of words learners can produce, disregarding repetition, and such different word forms are called types. The ratio of these two counting units, identified as type-token ratio (TTR), is used to estimate a text’s richness. However, this poses problems with texts of differing length (e.g. Daller and Phelan, 2013).

The regularities in languages, however, call for other types of counting units: lemmas and word families. In English, the regular past tense is formed by adding -ed and the regular comparative is formed by adding -er. Therefore, as Nation (2001, 2013) and Milton (2009) argue, once learners mastered these rules, the learning burden or difficulty of recognition, and even disponibility for production of moved and stronger, for example, become negligible. Although, there are also irregularities in a language and irregular forms such as
caught, better, or oxen might be more problematic for learners, it is necessary to have a systematic way of classifying these related forms.

*Lemmas*, then, are defined as the headword or stem and its inflexions and reduced forms (e.g. do, did, done, does, don’t, doesn’t, does not, did not, do not), usually belonging to the same POS. According to most researchers, lemmas seem to be the most reliable counting units (e.g. Vermeer, 2004). There are economical, pragmatic, and pedagogical reasons for this as well. Counting lemmas largely reduces the number of words in a corpus (Nation, 2013), therefore, it becomes more manageable to take samples and design tests; and learners at low proficiency levels are more likely to have mastered the most frequent forms. Thus, lemmas might represent the best counting unit in terms of learning burden (Nation, 2001; Milton, 2009; Schmitt, 2010). Psycholinguistic research also points to the fact that lexical entries are plausible to be stored in the ML as base forms and different lexemes are grammatically manipulated by adding regular inflexions (Aitchison, 2003; see section 2.3 for competing models). This also confirms the rationale behind adopting lemmas as counting units (e.g. Kremmel and Schmitt, 2016; Kremmel, 2016).

Another convention that comprises a broader category of word forms is *word family*. This includes not just the headword and its common inflexions, but closely related derivations as well (e.g. anchor, anchored, anchoring, anchors, anchorage, unanchored). As it comprises a larger number of words, it reduces word count even more than lemmas, albeit not without problems (Milton, 2009). From a counting and testing perspective, it is not so straightforward to decide which forms to include. Forms like anchored and anchors might not exert any problems to learners, however, anchorage or unanchored might. In some cases, many derived forms might even be problematic to NSs. Bauer and Nation (1993) define six levels of English affixes that can be utilised to decide which ones to include based on the learners’ age and proficiency. This is of significance from the perspective of this thesis as opting for either lemma or word family lists substantially influences the test-development
Chapter 2: Literature Review

criteria, which Chapter 3 will exemplify.

The concept of words is more complex, however, than just units of counting in a language, corpora, or the vocabulary size of NSs or NNSs. Words can be further categorised, for example, by frequency (see Chapter 3), POS, content or function, cognate status, abstractness and so on. Studies show that these factors have a significant effect in various contexts. Each of these factors can influence lexical recognition or retrieval. For instance, De Groot and Van Hell (2005), report that recall scores for L2 learners are 11% to 25% higher for concrete words, compared to abstract ones. Although it is out of the current perspective to address each of these factors and their effect, for TLA, the cognate status of words deserves special attention and will be discussed in section 2.3.

Even these relatively simple distinctions between words as single words or formulaic sequences, or units of count point to the complexity of the second facet, which is word knowledge. This is due to the fact that although LL appears to be gradual in nature, some forms of knowledge develops earlier than others (Milton, 2009; Schmitt, 2010) and this does not occur in a straightforward, linear fashion. Therefore, tapping into the different types of lexical knowledge at different stages (e.g. age, proficiency) can pose challenges. The following section thus considers vocabulary knowledge.

2.2.2 Dimensions of lexical knowledge

Vocabulary knowledge, due to its many aspects and complexity, has been defined and researched in several ways. In a language testing paradigm, it is just as important to understand what type of words we are testing, as defining what type of knowledge we study. Milton and Fitzpatrick distinguish between three different approaches in an attempt to characterise vocabulary knowledge: component, developmental (based on Read, 2000) and metaphorical (2014: 9).

In their view, the component approach entails the intricacies of different aspects of word knowledge, including the contrasts between these aspects. Thus, this category
comprises seemingly simple conceptualisations of word knowledge, such as the distinction between spoken and written forms of words and their meaning, the contrasts between productive and receptive vocabulary knowledge, to more complex and comprehensive lists of different aspects of word knowledge, such as Nation’s (2001: 27) ‘What is involved in knowing a word?’ taxonomy.

In educational settings and international assessments, such as IELTS, stakeholders tend to differentiate between four skills: listening, speaking, reading, and writing. These four skills, then, from a vocabulary related perspective, are mapped on to the false dichotomy of receptive and productive types of knowledge. The distinction between receptive and productive, sometimes also termed passive and active skills, is considered a false dichotomy as they are neither mutually exclusive nor jointly exhaustive. In other words, when listening or reading, which are receptive skills, we also produce meaning, that is, we comprehend and reinterpret what is being heard or read (Nation, 2013). The productive knowledge entails the acts of writing and speaking, or in other terms, producing language to convey meaning. Although Schmitt (2014) argues that this dichotomous distinction has great ecological validity, he also admits the exact relationship between the two is not absolutely clear.

From a testing perspective, Schmitt (2010) makes a distinction between meaning recognition and recall, and form recognition and recall. He links these terms to what he calls usage-based skills, that is, productive and receptive mastery of word knowledge (2010: 88). Recognition, in this sense, means that learners are able to recognise, or are familiar with, either the form or the meaning of a word from a set of stimulus words, for example. This is connected to receptive knowledge. Recall, however, which refers to learners being able to recall (activate or use) and produce either the form or a meaning, given a set of stimulus. This is related to productive knowledge. From a context-based perspective, Read (2000) contrasts recognition and recall with comprehension and use. According to this distinction, comprehension refers to testing vocabulary knowledge and understanding after students read
or listened to a text, and this links up with receptive knowledge. Use, which is connected to productive knowledge, refers to how well students can employ the lexis that is at their disposal or previously encountered in a text or elicited in a task.

In reading and speaking, students need to activate or recall target words and their meaning to facilitate comprehension or communication. This process is called lexical retrieval. While during reading students have to retrieve the corresponding meaning from their ML, in picture or word naming tasks, or during speaking, they are required to retrieve the corresponding form. This is important as vocabulary recognition tests with a multiple-choice format attempt to facilitate meaning retrieval given a set of target words or prompts. However, by providing a number of choices the task becomes incongruent with real-life situations, such as reading (Nation and Webb, 2011). In other words, it would be unrealistic to assume that during word recognition, participants also retrieve the corresponding meaning and just then select the appropriate meaning from the options provided. Nevertheless, considering the importance of the form-meaning link, tests such as the VLT and VST can provide a relatively good indication of the number of words students can correctly recognise (see Chapter 3 for the validity and limitations of meaning-recognition tests).

Nation (2001; 2013: 49) follows this receptive and productive dichotomy and enlists 18 different aspects of word knowledge, separated into the three large categories of knowledge: form, meaning, and use. These three broad varieties of knowledge are then broken further into their respective subcategories, making it the most comprehensive list to date. As Milton and Fitzpatrick (2014) argue, however, this level of complexification raises the issue of adequately testing each type of aspect in a capture-all lexical test. Furthermore, Schmitt (2010) argues that vocabulary learning is incremental in nature and, within this, the acquisition of knowledge of different aspects of individual lexemes is incremental as well. Therefore, it is difficult to see in what way the different receptive and productive aspects of word knowledge develop and how can tests explore this growth or development.
While Nation’s taxonomy and the receptive/productive word knowledge generally focus on individual words that the learner knows, and the details of that knowledge in terms of meaning and use, another frequently used dimension contrasts between *breadth* and *depth* of vocabulary knowledge (Anderson and Freebody, 1981). Breadth of knowledge refers to the number of words that are stored in the ML and depth encapsulates what is known about these words.

In a number of publications, Meara (1990; 1996; 1997) emphasises that he sees lexical knowledge in terms of links between different words, which can be associational or grammatical. In this view, depth of vocabulary involves increasing the number of links between different words in the lexicon, which in turn, affects the breadth of vocabulary as more lexis is acquired. Studies attempting to examine the development of breadth and depth of vocabulary knowledge, or receptive and productive for that matter, have been inconclusive, though (e.g. Nation, 2013). In a seminal paper, Schmitt (2014) arrives at the conclusion that the distinction between depth and breadth depends on how one conceptualises and measures the notions, as they can depend on the L1, proficiency in the L2, and test formats or items used. It is generally accepted though that learners can normally recognise more words than they can use productively, and this also entails that the difference between size and depth is less noticeable at low proficiencies (Schmitt, 2013). This also implies that depth of vocabulary develops at a relatively slower rate than breadth of vocabulary (e.g. Milton, 2009; Nation, 2013; Schmitt, 2010; 2013; 2014).

Based on Meara’s (1996) view on the links between words and that the more links are attached to a word, the easier is to access it, Daller, Milton, and Treffers-Daller (2007) create a three-dimensional lexical space, where breadth and depth are in contrast with *fluency*. Essentially, fluency refers to the ease, accuracy, and speed with which words are recalled and produced by learners. The ways knowledge progresses in this lexical space, for example, and the growth of breadth and depth of vocabulary belongs to the developmental approach.
Chapter 2: Literature Review

(Milton and Fitzpatrick, 2014).

The development of breadth of vocabulary precedes depth, as learners need to acquire a certain number of words before a more accurate or associational mastery occurs, and certain links are created between individual items (Milton and Fitzpatrick, 2014). Furthermore, receptive and productive mastery are seen as two extremes of a continuum, or a scale of the same type of knowledge, where these represent qualitatively different aspects (Melka, 1997). Others, such as Corson (1995), consider productive vocabulary as an integral part of receptive knowledge, which also includes partly-known or low-frequency words (of Greco-Latin origin) where the meaning is opaque, and words that are deliberately avoided (e.g. swear words or internet expressions, such as totes amazesh).

The third perspective to vocabulary knowledge is the metaphorical approach (Milton and Fitzpatrick, 2014), which consists of capturing the abstract nature of lexical knowledge through metaphors, such as the lexical space presented in a cubic form, where breadth, depth, and fluency can develop separately, but in reality they are likely to develop gradually and in tandem (Daller et. al., 2007) or a web of words (Aitchison, 2003), in which links are created between different words and the more links exist between them, the better the knowledge is.

These different dimensions of lexical knowledge shed some light on the complexity that is involved in knowing a word, the difficulties involved for learners, and the challenges teachers and researchers face in creating tests to investigate the lexicon. This multi-faceted nature of lexical knowledge is furthered in multilingualism for a number of reasons, which is going to be a leitmotif of the present thesis, and is oft-omitted from SLA and vocabulary studies. For instance, are learners able to use a word productively because of the efficient teaching they received, its frequency in the TL, or because it is a common cognate in their previously learnt languages? In this case, the lexical test would not be testing vocabulary knowledge or uptake, but cognate recognition skills (Nation, 2013). Thus, it becomes essential to consider not just lexical knowledge in the case of multilinguals, but the
relationship between their background languages, their cumulative LL experience or competence, and how these languages are stored, organised, and accessed. The following section, therefore, will discuss this lexical storage and organisation that is most often referred to as the mental lexicon from a multilingual perspective, while lexical access is discussed in section 2.6.

### 2.2.3 The mental lexicon (ML)

The ML, as the collection of words and the information regarding these in terms of meaning, pronunciation, syntactic characteristics and so on that an individual knows, can be conceptualised and described as a three-dimensional model, as proposed by Meara (e.g. 1996, 2009). According to him, the three dimensions are size, that is the number of words the learner knows; organisation, the structure or link between these words; and accessibility, how easily and fast these words and their information (orthographic, phonologic, or semantic) can be accessed. This is important as, from a storage perspective, lexical tests aim to quantify how many items the ML contains and the amount and degree of information that is known about them. Due to the intricacies of the ML and the various aspects of word knowledge, especially in the case of multilinguals, several definitions and models have been developed, as this section aims to demonstrate.

Regarding the organisation of the ML, Meara (1996; 1997; 2006) and colleagues (Wilks and Meara, 2002; Wilks et al., 2005) view it as a network in which there is a certain link between words (termed nodes) that get co-activated during language use in a certain direction. Relying on Graph Theory and later Random Autonomous Boolean Network Models, these models try to capture the mechanism behind the ML by simplifying it and ‘removing all the unnecessary complications of real life’ (Meara, 2006: 625). In brief, Meara turns the receptive and productive distinction into activated and unactivated binary functions and the model, then, consists of words that are linked to only two other words in the network. Based on input, these words can only be activated if at least one of their connected nodes are
activated or both. As the number of activations increase, the interactions between words increase as well. According to them, recently mastered words that entered the network have fewer connections and, therefore, words that were acquired at an early stage will likely be activated more often (e.g. Wilks et al., 2005). This could depend on the context, however. Although some words that are well established in the lexicon, and are frequently used productively, will be potentially activated more often. Thus, it can be argued that a recently mastered specialism for a biology student, for example, will quickly become part of the network and be used more frequently than common terms employed previously.

While the above refers to lexical activation or retrieval in the case of one language, here the interest falls on the mechanisms that describe such connections in the case of multiple languages. In relation to multilingual models, Meara (2006) adopts the idea of entanglement. According to this, some L2 words are entangled with the L1 network, but L2 words will also form a network of their own. When an L3 is added, the model becomes more complex, and in the one he portrays, the L2 will have links to the L1 and similarly, L3 will also have links with the L1, while not ruling out L2-L3 interaction in different models. What the models suggest is that activation in the L3 can have an effect on the L2 and L1 as well. Meara’s aim with these models was to highlight the underlying complexity that characterises the ML, and as he argues, general characteristics are observable even through these simplified networks. He claims these allow examining some fundamental issues by preserving only key features of the network. One benefit of these models is that based on the underlying assumptions, they produce stable results that can be interpreted reliably. These initial assumptions, however, create perhaps an indefinite amount of scenarios and as others have commented (e.g. Laufer, 2006), it is difficult to connect these to real-life communicative situations. Of relevance here is the fact that, as Meara proposes, different languages in the lexicon are not independent, links between the known languages are formed and, as a consequence, different languages will influence the acquisition or addition of new
languages to different degrees, depending on the similarity between them. This aspect will be further developed and elaborated on in the following models and empirical studies presented.

Alternatively to this network-view that Meara (e.g. 1990, 2006) and Aitchison (2003), inter alia, propose, ‘[w]e can conceive of it as similar to a printed dictionary, that is, as consisting of pairings of meanings with sound representations […] and definitions in terms of other words’ (Fay and Cutler, 1977: 509). Although this rules out entries that have no meaning attached to them or information on pronunciation (and spelling), it certainly points towards the direction of the ML being a system of form-meaning links. Therefore, the need arises to be able to establish the intricate nature of these form-meaning links in the case of multilinguals.

Szubko-Sitarek (2015) also recognises this similarity between a traditional dictionary and the ML. According to her, these are both built on frequency information and are organised along principles that characterise words. Nowadays, online dictionaries contain pronunciation and context as well, similarly to the ML. In the ML, however, this level of linguistic information is far more sophisticated, as speakers have to be aware of the context or register that a particular word fits, verbal and non-verbal features of the word, and in some cases, they have to overcome its inaccessibility, purposefully inhibit it, and find alternative word candidates as replacement. Another notable distinction is that the ML is more dynamic than a dictionary. While the dictionary is a collection of words and meanings, entries in the ML constantly change their depth, as new meanings, contexts or uses are added, new links are established within the given network, and rarely utilised words are also liable to become inaccessible (Szubko-Sitarek, 2015).

Traditionally, in psycholinguistics there are two conflicting approaches regarding the issue of representation of meaning in the lexicon. The so-called one-level or network model, proposed by Levelt (1989), which considers that meaning and conceptual knowledge about
the word are identical. The other approach, suggested by Bierwisch and Shreuder (1992), considers that a lexical entry’s meaning is composed of more primitive elements, and thus, the semantic properties of a word and the conceptual knowledge it refers to are separate. The common elements of these two approaches are probably the features of the internal structure of a word, which include: semantic form (meaning), grammatical form (morphology), syntactic from (the argument structure) and phonetic form (how it sounds or is pronounced).

Levelt’s (1989) model also makes a distinction between intrinsic and associative links. The formal entails words being connected by semantic, syntactic, morphologic, and phonologic properties. Thus, intrinsic links that are connected through meaning form a semantic field in which words are linked to their hyperonyms or co-hyperonyms (e.g. dog – animal or dog – cat), synonyms (big – large) or antonyms (up – down) and so on (Szubko-Sitarek, 2015). Evidence for these links mainly spring from error analysis of speech (e.g. Fay and Cutler, 1977) in form of malapropisms (slips of the tongue: saying word instead of world). Besides these formal connections, words can also be linked associatively (e.g. Aitchison, 2012; Fitzpatrick, 2007).

Apart from the form-meaning distinction and their storage, psycholinguists also tried to determine whether words themselves are stored as whole units or lexical entries, or as separate headwords and affixes (morphemes). The two major hypotheses that have emerged from these enquiries are the Full Listing Hypothesis (Butterworth, 1983) and the Decompositional Hypothesis (see Levelt, 1989). The first hypothesis postulates that each word form and its derivations are stored separately, in independent entries (e.g. play and player) and this has the advantage of access-efficiency as, for comprehension and use, they are accessed separately and, therefore, more rapidly (e.g. Vigliocco and Hartsuiker, 2005). On the other hand, the supporters of the Decompositional Hypothesis corroborate that morphemes (headwords and their inflexions and derivations) are stored in bundles (lexical entries) for reasons of economic storage (Levelt, 1989) and then decomposed into their
smallest units. Thus, the word *regenerate* will be ‘stripped off’ of the prefix and its meaning would be created from *generate* (meaningful unit) + *re-* (suffix). This distinction, in actuality, has particular relevance to lexical tests that are either based on lemmas or word families as units of counting. The reasoning behind this is that due to the regularities in a language, such tests assume that the most common inflexions and derivations of a word will be known.

The aforementioned examples of theories and models in psycholinguistics highlight only a small number of issues regarding the conceptualisations of the ML, how it is organised, the internal structure of a lexical entry, and how meaning is represented and accessed. These issues are further complicated in multilinguality. The major debates around the multilingual lexicon concern the dilemma of a single or multiple lexicons, and language-selective versus non-selective lexical access. Evidence in support of these views come from all type of experimental studies, mainly because ‘life experience offers ample support for each alternative’ (Kroll and Tokowicz, 2005: 531). The main reason for this is, as Szubko-Sitarek (2015) also argues, speakers of multiple languages can use their languages independently, however, codeswitching and transfer are common occurrences as well.

The fact that multilinguals are able to use languages independently supports the one-store hypothesis, i.e. there exists an integrated memory system comprising separate language systems. By contrast, the multiple-memory theory (two-store hypothesis, e.g. Van Hell and De Groot, 1998; Szubko-Sitarek, 2015) postulates that words in different languages are basically stored separately from a common conceptual store. It has to be noted that this dichotomy is characterised by task dependency – free recall tasks and studies on transfer support the one-store hypothesis, whereas comprehension, association, and lexical decision tasks are in favour of the two-store hypothesis. In simple terms, this postulates that there is a separate store for each conceptual, or semantic system, which serves each language (e.g. Paradis, 2004; Szubko-Sitarek, 2015).
Chapter 2: Literature Review

In order to incorporate the conceptual representation (meaning) of lexical items as a separate system, Paradis (e.g. 2004) proposes the three-store hypothesis. Two or more languages, according to Paradis, are connected to the same conceptual store in different ways. By these different ways in truth he means that while lexical representations from two separate languages are connected to one conceptual store, in each language the shared meanings can have (and often do) distinguishable or additional features attached to them (Paradis, 2004). The question, then, is: how and to what degree is the multilingual lexicon interconnected on the level of words and their meaning? The features of this interconnection between words in two or more separate languages, and their meanings are affected by a number of different factors, however.

On the one hand, there is the language user characterised by proficiency, age, learning experience, and the environment. On the other hand, there is the type of words defined by frequency and cognate status for example (Dijkstra and Van Heuven, 2002). Arguably then, not all languages can be accounted for in the same way in the ML and when examining multilinguals, the linguistic relationship between the known languages and user-related factors have to be considered.

This is congruent with De Bot and colleagues’ (2007) argument that all variables are interconnected in the system and thus, changes in one variable necessitate a change in the whole system – called the Dynamic Systems Theory. They also question the common assumption that all L1 systems are the same and depending on the acquisition, multilinguals might be having different systems, number of subsystems, and modes of access (De Bot et al., 2007). Evidence for this is found in Cieslicka (2000), for example, who examined L2 learners based on their proficiency and learning strategies. Her findings are that L2 proficiency and learning strategies influence the pattern of lexical representation and how words are processed, and therefore, she advocates that the L2 context of acquisition should be controlled for in bilingual studies.
In order to account for such differences of language mastery in the case of an L1 and L2, and the nature of connections between words and concepts, Kroll and Stewart (1994) propose *The Revised Hierarchical Model*. According to this model, there is a shared conceptual representation between two independent lexical representations for each language, where the L1 is assumed to be larger than the L2. Consequently, the strength of L1-L2 connections and the conceptual storage differs here again. Moreover, as concepts develop together with the L1, the connection between the two will be considerably stronger than semantic links between the L2 and concepts (e.g. Clenton, 2015; Kroll and Stewart, 1994). Similarly, L2 links will be stronger towards the L1 than backwards. Evidence for this can be found in translation studies, confirming that forward translations (L1 from L2) are typically faster than from the mother tongue to the FL (e.g. Kroll and Dijkstra, 2002; Kroll and Stewart, 1994). Critics of this model suggest that L2 lexico-semantic mappings may not yet be developed in early L2 acquisition (Duyck and Brysbaert, 2004) and the substantial proof of an integrated lexicon (e.g. Brysbaert & Duyck, 2010).

The idea of an integrated or at least heavily interconnected ML appears as having more advocates when not just an L1 and L2, but more languages are involved (L3/Ln). According to Gabrys-Barker (2005), who adapted the *Revised Hierarchical Model* to L3, items in the multilingual memory are interconnected by lexical and conceptual links. Both types of links can exist within the same language or across languages, where the lexical links represent the form or translation, and the conceptual links relate to the semantic field. Words in this integrated memory are accessed via either link according to task, stimulus, language proficiency and competence, or language dominance.

Another confounding variable to the ML is the addition or integration of new lexical items. Through the *Parasitic Model*, Hall and colleagues (e.g. Hall, 2002; Hall and Ecke, 2003) hypothesise that new lexical items will, where possible due to similarities, integrate with the rest of the already existing network ‘with the least possible redundancy and as
rapidly as possible’ (Szubko-Sitarek, 2015: 82). Were this true, in the case of multilinguals if an L1 meaning is attached to non-cognates in the Ln, due to the cognate facilitation effect, the meaning of a new word would become parasitically attached to the language (L1 or L2) that has an already known cognate in that language. Therefore, it is amenable to theorisation that some L3 words would have an initial form-meaning link with the L1 yet others would be connected to the L2 or Ln, depending on cognate status and language dominance.

To illustrate with an example, a Hungarian NS from Romania with English L3 has to study another FL during their BA. In most cases, this is either German or French. A learner of German, for example, encounters the following new items:

**Example 1. Cognates in English and German but not in Hungarian or Romanian**

a. *bring* – *bringen* (German - GE)
b. *good night* – *gute Nacht* (GE)
c. *mine* – *mein* (GE)
d. *apple* – *Apfel* (GE)

**Example 2. Cognates in two or more different languages**

a. *wine* – *Wein* (GE) – *vin* (Romanian - RO)
b. *post* – *Post* (GE) – *poștă* (RO) – *posta* (Hungarian - HU)
c. *potato* – *Kartoffel* (GE) – *cartofi* (RO)
d. *screw* – *Schraube* (GE) – *șurub* (RO)

Since the words in Example 1 are non-cognates concurrently in Hungarian and Romanian, attaching an L1 or L2 link to these would require more cognitive effort from the learner than linking it to English. However, in the case of Example 2 the form-meaning link to L3 English is not that obvious. *Wein* is cognate with *wine* and *vin*. Similarly, the German word *Post* can be linked to either the L1, L2, or L3. In these cases, it can be hypothesised that the link will be made either to the more dominant language, or the one with the most recency or frequency of use. The last two German words, *Kartoffel* and *Schraube*, which are
cognates only in Romanian, however, can only be linked to the L2.

Based on their previous research, Hall et al. (2009) further argue that the parasitic attachment creates a link between the short-term memory, where the TL item is stored, and the long-term memory, where previously known languages are stored. They also posit that the parasitic attachment to the long-term memory can be done on three different levels (*form* or *lexeme* in Levelt’s (1989) framework, i.e. orthographical or phonological representation; *frame* or *lemma*, i.e. syntactic category and sentential deployment; and *concept*, i.e. meaning), at any level and to any of the languages where similarity exists. Conversely, non-cognates in Hall’s model may only create a frame (or grammatical) attachment initially, then a conceptual one, resulting in possible errors such as ‘*Acapulco likes me for I like Acapulco*’ (Hall et al., 2009: 162) from the grammatical link between *like* and *gustar* (SP). They also extrapolate that the parasitic connections are governed by psychotypology, where the proximity of Romance languages to English can be more significant than English and German similarities.

This integration mechanism governed by language proximity or psychotypology has been observed by others as well (e.g. Herwig, 2001; Singleton, 2002). Herwig (2001) describes a single ML, consisting of dynamically interacting subsystems that are governed by the typological distance between the languages, the learners’ perceived linguistic distance, language proficiency, and LL history. Singleton (2002) depicts this psychotypology between languages in terms of different stages. Accordingly, he conceptualises that in the initial stages of acquisition, the L3 is a subordinate system of the typologically closer language, later they share only concepts, and at highly proficient levels, the structure of the ML comprises coordinate systems where both lexical and semantic representations are separate.

Early theories in psycholinguistics assumed that the separate or integrated storage was also in line with the way the languages are accessed in the lexicon. More recent studies (e.g.
De Groot, 2011) nevertheless reveal that a separate storage does not necessarily imply *selective access* (i.e. during a task, languages are accessed in the lexicon separately), but may be a result of different activation patterns (i.e. task demands) (Szubko-Sitarek, 2015). The *non-selective access model*, on the other hand, entails simultaneous activation of languages and this has perhaps found a wider empirical support in eye-tracking and brain-imaging studies (see Szubko-Sitarek, 2015). Studies similar to Dijkstra (2005) confirm that in the case of interlingual neighbours (similar cognates) candidates in both languages are activated and this is valid not just in the visual, but in aural experiments as well. Concurrently, based on reaction time studies (e.g. Lemhöfer et al., 2004), it has been argued that if language access was selective, then the cognateness of a word would not affect reaction times. It has to be noted that although the non-selective access model has gained wider empirical support, researchers have also raised concern about the relationship between findings and task dependency, emphasising that context has an influence on CLIs (Szubko-Sitarek, 2015).

Finally, as Szubko-Sitarek (2015) points out, apart from the above-mentioned properties of the lexicon, one additional underlying, and much-debated problem, is the size of the multilingual ML. This has not been addressed in depth by former research. Since the aim of this thesis is to explore the size of the lexicon in the case of multiple languages and how two proximate languages are integrated and influence each other within the ML, the following section focuses on cross-linguistic similarities and influences at the lexical level.

### 2.3 Cross-linguistic similarities: cognates in vocabulary research

Cross-linguistic similarities can be defined as the common structural (phonologic, morphologic, grammatical, semantic etc.) and lexical similarities amongst languages (Otwinowska, 2016). For the current purposes, the emphasis will be on lexical similarities as the most salient characteristic of the degrees of congruence between the known languages (De Angelis, 2007; Jarvis and Pavlenko, 2007). Lexical similarities among languages are
due to the fact these can share a common ancestor, such as Latin, and that throughout the years, certain languages also influenced each other. For instance, French as the language of European aristocracy or English as a global language and the language of modern science, business, and education. Consequently, many languages share a considerable number of words that are very similar or even identical, called inter-lingual neighbours or cognates. The term cognate derives from the Latin word *cognatus*, meaning blood relative, and in historical and contact linguistics this connotation is restricted purely to a common etymological origin or defined as ‘words in different languages which have descended from a common parent word’ (Schmitt, 1997: 209). This definition is rather exclusive and limited though.

In psycholinguistics, the definition of cognates has been extended to include other shared words between languages, such as borrowings and internationalisms (see Helms-Park and Dronjic, 2016). Thus, cognates are regarded as orthographically and/or phonologically similar translation equivalents, or as De Bot puts it, ‘words with similar form and meanings in two languages’ (2004: 19). The reason why etymological perspectives are disregarded from a LA point of view is that language learners and users will not necessarily be aware of the difference between cognates that share a common ancestor, loanwords, and internationalisms, so essentially, they can be treated equally (Jarvis, 2009; Otwinowska, 2016). Broadly speaking, thus, cognates can be found not just in historically related languages, but in typologically more distant languages as well, such as Chinese, Japanese, or Polish in relation to English.

On this account, the definition widely used in SLA and psycholinguistics will be applied, disregarding etymological and historical perspectives. Thus, cognates, or interlingual neighbours, represent words that are orthographically identical (e.g. *sport*; identical cognate in EN, RO, and HU among many) or similar in two or more different languages (e.g. *calculate* or *invite* - EN; *calcula* or *invita* - RO) and they share a common
meaning (e.g. Dijkstra, 2005, Molnar 2008). However, this can also depend on the context. Conversely, words that are identical or similar cross-linguistically, but there is no semantic overlap, are called interlingual homographs or homophones (De Bot, 2004), or false cognates (commonly: false friends). For instance, *avertisment* (RO) and *advertencia* (SP) closely resemble the English word *advertisement*, but in fact, they mean *warning* or *sanction*. Similarly, *confectionary* is very close to *confeționare* (RO) and *confeccionar* (SP), albeit they mean *to fabricate*. These words, can represent deceptive items for learners, which can result in miscomprehension or production errors.

This potential deceptive nature of cognates caused language educators in the past to focus on false friends and completely mistrust cognates, and researchers to explore the negative effects of cognates, especially in the field of error analysis (Granger, 1993). Others see this preoccupation with false friends as ‘stultifying and negative’, and promote the idea of deploying cognates as a positive source for retention (Pons-Ridler, 1984: 89). This facilitative effect of cognates has been supported by statistical evidence subsequently and a wide range of studies that promote cognates for LA.

Meara (1993) estimates the number of English and Spanish cognates as 3,000 and the number of French and English identical cognates as 6,500 and non-identical as much as 17,000. The special status of English in the Germanic and Romance families has been confirmed by Schepens et al. (2012). They utilised computational tools developed in artificial intelligence and looked at orthographic similarity distributions of cognates across English, German, French, Spanish, Italian, and Dutch. Their conclusions indicate that English shares the most cognates with French (9,286), followed by Dutch (8,609), Spanish (7,837), and then German (7,750). Additionally, English is the language that shares the most cognates with all other languages. Although they relied on a translation database which, however large, is considered a semi-complete lexicon of these languages, this can be regarded as an accurate approximation of the typological distance between these languages.
and as evidence for the close relationship between English and Romance languages.

In relation to the proportion of cognates and false cognates between Spanish and English, Morán-Molina (2010) points out that the General Service List consists of 32% of cognates, whereas the Academic Word List (AWL) comprises 71% of cognates in comparison to only 7% of false cognates. Overall, he argues that there are over 20,000 real cognates compared to only 200+ false cognates between English and Spanish. Hammer (1976) approximates the ratio of English-French cognates to deceptive cognates as 11,000 to 950, and intimates that on the one hand, studying cognates could largely reduce the memory effort required to master vocabulary, and on the other hand, using false cognates incorrectly is not the most enduring type of error when it comes to LA.

The aforementioned relationship between English and Romance languages and the proportion of cognates and false friends in English are important since Romanian and Spanish belong to the same Romance family, the number of English cognates that Romanian L2 speakers have at their disposal will be a considerable amount from a language testing perspective. As a consequence, the proportion and distribution of cognates is a core factor in the studies featured herein and, as is contended throughout the thesis, the presence of cognates should be accounted for when it comes to vocabulary tests.

Due to their similarities in proximate languages, cognates have been shown to have a large effect on SLA and teaching (e.g. Carlo et al., 2004; Dijkstra et al., 2010). This is most commonly referred to as the cognate facilitation effect emphasising that cognates can aid LL, comprehension, and can be used effectively in teaching. Jessner (1999), for example, proposes that already known languages should be used in EFL classrooms so that students can draw meaningful comparisons. On a similar note, Carlo et al. (2004) corroborate the benefits of enabling students to activate their cognate knowledge as this can enhance the comprehension of unknown English words.

Congruent with these suggestions are the findings in different areas of
psycholinguistics that suggest, employing different experimental paradigms, that cognates are recognised faster than non-cognates in both visual (e.g. Lemhöfer et al., 2008) and auditory modalities (Marian and Spivey, 2003). Concurrently, similar cognate facilitatory effects have been found in word production (e.g. Kroll and Stewart, 1994); learners using different scripts (e.g. Kim and Davis, 2003) and the effect is considered to be stronger if the stimulus words are cognates in more than two languages (Lemhöfer, Dijkstra & Michel, 2004). According to Dijkstra et al. (2010: 4) ‘[t]he cognate facilitation effect has often been taken as evidence for a bilingual lexicon that stores words of two or more languages in an integrated fashion and/or for a lexical access procedure that activates word candidates in several languages in parallel’. Furthermore, they affirm that the degree of form-similarity and frequency of cognates are influential determiners of this effect. Ergo, cognates can inform on how words are stored and retrieved from the ML and as such, how they influence lexical knowledge and, from the point of view of this thesis, how they contribute to and influence VS.

It is clear though that cognates in different languages will have slight variations either on a formal (orthographic and/or phonologic) or on a semantic level (e.g. a polysemous word in a Romance language might have only one or a much-specialised meaning in English), or both (Otwinowska, 2016; Szubko-Sitarek, 2015). Therefore, cross-linguistic lexical similarity is seen on a continuum, where cognates range from identical cognates (there is a complete morphological and semantic overlap), partial cognates (which includes regular and irregular morphological differences with only a partial semantic overlap) to deceptive or false cognates (where words share a formal similarity, but no semantic overlap) (Jarvis, 2009; Ringbom, 2007).

It is worth noting that Latin words exist in Romance languages as a direct descent. Therefore, the meanings of Latinate words might represent a simple concept and such words are more frequent for a Romance speaker than an English one. Due to the development of
English and the way words entered the language, often through French because of the level of their sophistication, cognates in English tend to be of lower frequency (Lubliner and Hiebert, 2011; Nagy et al., 1993). Researchers suggest that the influence of Latin on English and Romance languages offers speakers of these languages a common linguistic heritage, which can be regarded as a potential fund of knowledge (Moll et al., 1992). In relation to academic words, Cummins (1994) argues that in English these are concept-laden words that are in many cases unique to a specialist field and knowledge of these concepts from an L1 will advance their acquisition.

Some studies have found that in spite of the potential advantage that cognates offer, it is not uncommon for bilingual students to fail to notice cognate words (August et al., 2005; Feldman and Healy, 1998; García, 1991; Nagy, 1995; Nagy et al., 1993). According to Lubliner and Hiebert (2011), it is not yet fully clear why cognate identification is problematic in some cases, whereas other studies suggest that it is an automatic process (see Hall et al., 2009). On the one hand, the reason for this could be that in the case of typologically more distant languages, a number of cognates and their degrees of similarity decreases, thus learners do not perceive the languages as mutually influential and this might stop them from searching for congruencies (Odlin, 1989). On the other hand, the number of cognates that learners recognise could be influenced by individual differences, cumulative LL experience, which might include explicit cognate instruction and other complex cognate characteristics (such as meaning, form or sound based) (August et al., 2005). Lubliner and Hiebert (2011) also consider that in order for students to recognise cognates in English, first they have to be familiar with the Spanish word and they have to be able to access its meaning based on the target word’s orthography and phonology.

Another potential feature that makes cognate recognition difficult for students is polysemy. Bilingual students, in many cases, showed knowledge of only one meaning of polysemous words (August et al., 2005). Nagy et al. (1993) also came to the conclusion that
middle school learners had difficulties with inferring meaning from context partially because of the number of unknown words.

From a didactical perspective, it is said that the number of cross-linguistic similarities between the background and TLs can be a decisive factor in defining how difficult will be to learn or understand the TL (Otwinowska, 2016). Since learning FLs is basically attaching or associating the unknown with the already known, according to Ringbom (2007), cognate vocabulary appears to be an ideal way to learn quickly and effortlessly. Similarly, learning a language that shows congruencies with the already known languages will be easier in comparison to more distant languages. Additionally, since (Indo-)European languages share the higher vocabulary of art, science, and abstract language (e.g. these words are cognates in not just Romance languages, but in Hungarian, Polish, and German [and probably many others] as well: rapid, music, soprano, abstract, academic, aesthetic, calligraphy, pedagogy, conceptual, symbiosis, trigonometry, rhapsody etc.) the required cognitive effort to master such lexis is considerably less (Odlin, 1989).

Unfortunately, even though the aforementioned suggests that cognates have the potential to constitute a valuable resource learners and, thus, bridge a gap between the known and the TL, evidence suggests that cognate recognition is far from a simplistic process, for it has been corroborated that learners do not automatically make use of such knowledge or not to the same extent. Another way to explain this is to inspect evidence that takes into account L2 learners’ proficiency, which includes language and metalinguistic awareness that can influence the extent to which learners can notice cross-linguistic similarities. The research on the exact relationship between proficiency and cognate facilitation is inconclusive, however. In his review, Serrander (2011) proposes that the facilitation is more noticeable at lower levels. As discussed later, Laufer and McLean (2016) came to the same conclusion as cognate recognition helped students with a basic proficiency more.

In contrast, studies such as Hancin-Bhatt and Nagy (1994) or Méndez Pérez, Peña and
Bedore (2010) argue that the amount of exposure positively influences cognate recognition, and, it seems, students with a higher proficiency and metalinguistic knowledge benefit more from the presence of cognates. A certain influence by proficiency is often observed, however, this might be task dependent or influenced by other underlying factors. Costa et al. (2000) investigated through a picture-naming task whether Catalan-Spanish bilinguals named cognate words in Spanish faster than non-cognates. The results indicated that cognates induced advantage in both languages, especially in the non-dominant language, which is an indication that the cognate facilitation effect is influenced by proficiency. By comparison, Brenders, Van Hell, and Dijkstra (2011) found that L2 cognates were named faster than non-cognates, albeit this was not the case in the L1. This suggests that cognates have a facilitative effect in the L1 only if the L2 knowledge has reached a certain threshold. Van Hell and Dijkstra (2002) came to a similar conclusion and suggest that for cognate facilitation, participants have to reach a certain FL level and the effect is negligible for L1.

Apart from proficiency, cognate recognition can also be influenced by age. García (1991) found that cognates were only recognised by young people after EI. Molnar’s (2008; 2010) findings also show that in the case of secondary school students, cognate recognition is automatic. Hancin-Bhatt and Nagy (1994) and Nagy et al. (1993) found that as learners become older or are exposed to the TL for longer, unavoidably proficiency and metalinguistic awareness increases as well, and therefore, their ability to recognise or capitalise on cognates in a reading task leads to better comprehension.

The aforementioned looked at general evidence on the effect of cognates and their effect in the case of FLs. As many researchers argue, cognates in general can contribute to LA and, by extension, considerably reduce the insurmountable number of lexemes learners need to master in order to function in a TL. The following section will critically synthesise studies that employed vocabulary tests similar to the VLT and the VST in order to explore the intricacies of prior lexical influence and the effect of cognates.
2.4 Vocabulary knowledge and the influence of background languages

Hitherto, the focus has been on the core elements that define and influence vocabulary knowledge and lexical tests. The preceding sections introduced key aspects that pertain to these issues, such as different ways of defining and categorising words, the complexities of lexical knowledge, and the ML. Additionally, section 2.3 addressed the issue of cognates in multilingualism. The second part of this chapter reviews empirical studies that explored the influence of prior lexical knowledge on VS.

In relation to the lexicon and the influence of cognates, our knowledge is still limited, especially in TLA. One notable example is the study conducted by Augustin Llach (2012), which investigated the influence of different L1s (German and Spanish) on English (L2) vocabulary. The study had two primary aims. On the one hand, it set out to explore beginner, primary school EFL learners’ lexical knowledge and the difference in their VS based on the L1. On the other hand, the researcher was interested in how word origin influences learners’ VS on the 2k level. The study involved 41 Spanish and German EFL learners respectively, with similar LL history and comparable levels of proficiency. This was based on a cloze test and a reading comprehension pre-tests. In order to measure VS, two different tests were administered, the 1000 Word Test (WT), which is a translation test with a similar format to the VLT (see Augustin Llach, 2012), and the 2k level of the VLT (Schmitt et al., 2001). The results of the study indicate that both language groups attained scores over 500 on the 1k level and below 200 on the 2k level, and the author suggests that, disregarding the L1, learners’ vocabulary knowledge is comparable in English and due to the similarity in scores, the L1 does not influence the acquisition of high-frequency words.

While this might hold true to some extent, whether the two tests are directly comparable is questionable at best. One clear problem with this comparison between the two groups and the two tests is that the paper does not specify whether the WT is frequency based and even if it is, which word list was utilised to obtain the items. The second one is that the
Vocabulary knowledge and the influence of background languages

WT, as a translation test, which includes function words (e.g. could, this, of, in order to, my) that are normally deemed as level 0 words and, thus, eliminated from lexical tests, makes the scores problematic to be compared to the VLT. The ratiocination for this is that, as Elgort (2013) asserts, with translation tests it becomes ambiguous whether learners are translating L1 meanings to the L2 target word (form recognition) or attempting to find the L2 meaning in the L1 (meaning recognition). Furthermore, forward translations represent less burden for the learner, and therefore, the fact that both translation and meaning are provided within the same test item makes it more manageable for the testees than the search for meaning in the L2. Congruous with this is the finding in Elgort (2013) as she also intimated that the bilingual version of the VST elicited significantly higher scores than the monolingual one. Consequently, the difference in the 1k and 2k scores in Augustin Llach’s study appears to be an effect of task demands rather than frequency.

Regarding the effects of cognates, Augustin Llach relies on Perez Basanta (2005) who claims that the VLT comprises a balanced number of words originating from Latin and German respectively, and thus, the test is appropriate for such an enquiry. According to her findings, there are 13 and 21 words of Greco-Latin origin and 17 and 9 of Anglo-Saxon on the 1k and 2k levels respectively. Hence, she argues that ‘[t]his clearly shows that as word frequency decreases, the test contains more words of Greco-Latin origin and thus, more cognates for Spanish learners. It seems reasonable to suggest that Spanish learners will show better results in the tests of lower frequency bands’ (2012: 9). An increased number of Latin cognates on lower frequency bands would not necessarily mean that they would outscore the other group, however.

The question that needs to be asked is how much advantage Latin cognates give to Spanish test-takers and how this can be accounted for? The results in her paper suggest that Anglo-Saxon words were better known by German learners and Latin words were correctly answered by a higher number of Spanish learners. However, as there are slightly more Latin-
origin words in the test, this discrepancy could be misleading. Additionally, the way cognates have been categorised might not be entirely accurate to establish the influence of the L1. The Anglo-Saxon or Greco-Latin words might be problematic for the target groups, as some Latin words can be less opaque for a French learner than a Spanish one (the EN trouble = problema in SP, le trouble in FR).

Despite these issues that have to be considered when interpreting the findings of this study, the questions asked are valuable indeed. The impact of different L1s can lead to varying vocabulary scores and, while researchers need to understand the intricacies of this, teachers might also find it relevant from a classroom perspective. The problem is, however, that when measuring VS using tests such as the VLT, the scores on one level only might not be as reliable as the test in its entirety due to the limited number of test items (e.g. Nation, 2013). Comparing translation and L2 form-meaning recognition leads to highly contrasting results, as this study also demonstrated that the differences are not necessarily due to frequency effects. Furthermore, the issue that the VLT was not designed to be able to distinguish between vocabulary knowledge in terms of form-meaning links and cognate recognition should be considered too. To decide whether the cognate is known because of the L1, in the case of young learners especially when cognate recognition is not automatic yet (Nagy et al., 1993), or simply because it is known in the L2 is quite problematic. Finally, the distribution of cognates on vocabulary tests is an under-researched area and so it should be based on the similarities between two languages (e.g. Spanish and English) and not word origin (Greco-Latin or Anglo-Saxon).

In order to investigate the effect of background languages on English vocabulary acquisition, Chen et al. (2012) tested the lexical knowledge of 244 school children and how this is influenced by cognate awareness in three different groups: Spanish-speaking English learners, Chinese-speaking English learners and English monolinguals, all attending Canadian schools. The study showed a clear contrast between the monolingual and the two
bilingual groups, however, the difference between the two bilingual groups was only significant in terms of cognates. The authors corroborate that the fact that cognate awareness in the case of Spanish-speaking learners reduced the gap between them and the monolingual group is a clear indication that cognates influence vocabulary development. Moreover, this suggests that even for learners of similar age and exposure to English, the typological distance between the languages will result in a measurable outcome in terms of vocabulary development due to the cognate facilitation effect. This effect can be extended to not just the L1, but in the case of multilinguals to Ln as well. In TLA, findings suggest that the facilitatory effect of cognates, or in other terms positive lexical transfer, will be typically from the (psycho)typologically closer language (e.g. Cenoz, 2000; Gass and Selinker, 2001; Kecskés and Papp, 2000).

In a study involving participants with a DLC more closely related to the one presented in the current thesis, Molnar (2008) investigated the lexical knowledge of three different groups. She compared English VS in the case of Hungarian monolinguals, Romanian monolinguals and Hungarian-Romanian bilinguals, and investigated the potential influence of participants’ L1/L2. Employing the VLT, she tested 200 \((RO = 60; HU = 43; HU-RO = 97)\) high school students with over six hours of English exposure (intensive classes). The findings confirmed Molnar’s (2008) original hypothesis: Romanian monolinguals outperformed Hungarians, and Hungarian bilinguals outperformed Hungarian monolinguals. These differences were significant on overall (cognate plus non-cognate) and cognate only scores as well. On non-cognate scores, however, the difference between the Hungarian monolingual and bilingual group was non-significant, a result which is in line with Chen et al. (2012). This indicates that bilinguals could capitalise on cognates, however, no difference could be ascribed to non-cognates.

Molnar furthermore argues that the difference between the groups becomes more accentuated beyond the 2000 levels on the test. This is, in my view, due to the fact that
knowledge on low-frequency levels decreases and the high number of cognates on the UWL level provides an advantage for Romanians according to their Romanian proficiency. This highlights the factuality that proficiency in the background languages, if they are close to the TL, may positively influence vocabulary scores, and therefore, accounting for these languages and cognate knowledge is important. Additionally, it also highlights the problems with cross-cultural comparisons in term of VS. If the VLT was an admission test and cognate scores were not controlled, the Hungarian monolingual group could be argued to have a disadvantage compared to their bilingual peers. Conversely, these results also exert that the effect of previously known languages and the proficiency in that language is more prominent than the bilingual skills. Molnar also corroborates from the findings that the reduced learning burden in the case of cognates might put the Romanian group in advantage by allowing them to concentrate on non-cognates and acquire them at a faster rate, which is supported by Morán-Molina (2010).

Overall, Molnar’s study is an excellent example of the fact that the exaptive VLT is a useful tool to compare vocabulary knowledge across cultures and borders if linguistic resources are accounted for. In the case of multilinguals, however, it is not just the TL and the level of acquisition that is important but the facilitative effect of background languages. One aspect that Molnar’s study fails to address is whether the test was differentiating between receptive lexical knowledge in English or cognate recognition skills. In comparison to Muñoz (2000), inter alia, where language competence in all background languages was tested, it becomes problematic to link the results to the interdependence hypothesis (see section 2.7), as the L1 and L2 proficiency of bilinguals were not tested. Monolingual studies in English show that there can be a great variation even in NS’s lexical knowledge thus, assuming L2 proficiency leads to tentative results only, a limitation addressed in Molnar (2010) and reviewed in the next section.

In order to evaluate the VLT from the point of view of cognates, Cobb (2000) turns to
Meara’s principle for a standard test and examines how well the test ‘answers questions about how many words people know, [...] and how these factors are related to other aspects of linguistic competence’ (1996: 41). Cobb selected the 2k and UWL levels of the VLT and used them as a placement test (in conjunction with a reading, error identification, and a writing task) for Canadian Francophone students entering university. Students tended to search more Anglo-Saxon words in the dictionary than Latinate ones. One interesting aspect of this study is that the items and glosses were rated for Greco-Latinness and Anglo-Saxonness, and following this, the author suggests that the 2k level balances well words from these two families, although with minor Greco-Latin bias.

The findings revealed that students were inclined towards correct answers on cognate items more than non-cognates. In other words, the more Greco-Latinate rating a word had, the higher its accuracy (or guessability) was. Cobb raises concern about students’ inability to answer Anglo-Saxon words correctly as the proportion of these is very high among the most frequent words and, therefore, this could result in weaknesses in other linguistic skills. To examine whether the exclusion of cognates would lead to a better correlation between vocabulary and other language skills, he tested 73 future ESL teachers. This resulted in a substantial increase in the correlation coefficient (from .6 to .9). However, it should be noted that although non-cognate VLT items were used from the same frequency band, the test itself has been changed to a C-test, which only partially relies on recognition (Cobb, 2000).

In a more recent study Helms-Park, Petrescu, and Dronjic (2009) examined the cognate facilitation effect in the case of Romanian and Vietnamese undergraduates of English in Canada who completed the newer version of the VLT by Schmitt et al. (2001) with an additional test that comprised very low-frequency English words. They observed the cognate facilitation effect in the scores of the Romanian learners, despite the groups were comparable on non-cognate items. On the 10k level, the difference between the two groups became more significant with a higher effect size both for cognate and non-cognate items. Regarding
uncommon English words, they report that although there was a significant difference between cognate and non-cognate responses for the Romanian group but not the Vietnamese or English NS group, the Romanians overall matched NSs. These findings reiterate the effect cognates can have on scoring and the importance of background languages.

So far, the focus has been on studies that mainly investigate the effect of cognates in the case of typologically close languages to English. The fact that cognates have an influence on the acquisition and testing of English as a TL, even if the background languages are not closely related on a lexical level, serves as further evidence that monolingual vocabulary tests that do not account for learners’ previous knowledge might over- or underestimate VS. The exact nature of this effect has been addressed by Laufer and McLean (2016) in the case of two typologically more distant background languages, namely Hebrew and Japanese. They inspected loanwords as a special type of inter-lingual neighbours and hypothesised that common loanwords are already part of students ‘potential vocabulary’ (Palmberg, 1987), and therefore, they do not have to be learnt. Thus, in the case of L2 vocabulary tests, as the items are randomly sampled from frequency bands without taking into account the L1, they cannot reflect the real proportion of cognates or loanwords, and therefore, such tests have the potential to provide misleading results. As Laufer and McLean argues, ‘the varying presence of loanwords or cognates for different L1 populations in vocabulary tests […] present a danger to the content validity of vocabulary tests, the interpretation of tests scores, and the research that uses vocabulary tests’ (2016: 204).

Consequently, their aim was to explore the effect of cognates in vocabulary tests and how the number of loanwords influences the scores in different test modalities (word form recall, form recognition, and meaning recall) and L1 groups. They used the item displacement method (i.e. the test is scored twice, with cognate items and when these are replaced with non-cognate items) to discover the effect of 6 Hebrew and 13 Japanese loanwords in the 80-item test.
The findings suggest that loanwords significantly increased test scores in all three testing modalities and for both language groups, even though the percentage of Hebrew loanwords was only 8% compared to 6% in Japanese. They found a significant difference for lower level students only, meaning that there is a probability that more advanced students were as familiar with the words that replaced the loanwords as with the loanwords themselves. The largest score increase (17% for Hebrew and 28% for Japanese) was found in the case of both L1 groups for students with basic L2 proficiency.

Based on this, they conclude that the cognate effect is most prominent in the case of beginner and intermediate learners as their limited vocabulary knowledge on low-frequency bands can be influenced most by the number of loanwords that belong to their potential vocabulary. This is in line with Elgort (2013), who also found that response accuracy to cognate items is most noticeable on lower frequency bands. Jordan (2012) also inferred that Japanese students translated more loanwords overall than non-cognates, albeit this difference was not true for items on the 1k level. Again, this is due to learners’ proficiency as if they are very familiar with high-frequency words, the items’ cognacy will be less salient or more items are needed to detect such differences.

Another important point from Laufer and McLean (2016) is the fact that, as the authors argue, VS tests are often administered to testees with a mixed proficiency. Therefore, a random sampling of items, which disregards cognates, undermines the test’s capability of accurately distinguishing between proficiency groups.

In sum, this study is a prominent example that the scores of a vocabulary test can be influenced by different L1s, test modality, and learner proficiency. Subsequently, to get a better picture of VS, the effect of cognates and learners’ background languages should be all accounted for. This is also one limitation of the present study as, although the authors describe the Japanese participants as Spanish-Russian majors, the inference from these languages is not taken into account and, also, there is no mention whether the Israeli students
could have an influence from other, more closely related languages to English. Since Spanish has a high proportion of cognates in English and, according to Elgort (2013), the VST comprises 34% of Russian-English cognates on average, it would have been valuable to see whether cognates from other L2s have an effect on the results.

Another limitation also has to be mentioned – the distribution of cognates on frequency bands. Vocabulary studies that employ random stratified sampling are at risk of under- or over-representing cognates, which, in turn, will provide inaccurate VS estimates. This has been investigated by Meara and colleagues.

Meara and Buxton (1987) infer that depending on the form of the pseudoword, it may be more attractive to speakers of a particular language family than others, such as the word ‘observement’, which might deceive Romance language speakers as a real word. However, it will be less attractive for a German speaker. Since hits to pseudowords are penalised by Yes/No tests, speakers of a certain L1 might score much less, depending on the particular pseudowords, thus risking VS underestimation or unstable results unless pseudowords and cognates are controlled for. This caused in studies, such as Meara and Jones (1990), that VS correlated to a lesser extent with other linguistic skills of Francophones compared to learners with different L1s.

Meara, Lightbown, and Halter (1994) explored the effect of increasing the number of cognates for Francophones and whether this leads to the overestimation of test-takers proficiency, or conversely, the exclusion of such items could lead to an underestimation. They concluded that in order to not compromise test validity, the number of cognates should reflect the proportion of cognates in real language, which is in line with later assertions by Eyckmans (2004) and Elgort (2013).

Another proposition for dealing with the effect of cognates and the influence of L1 is provided by Meara (1992). He argues that the cognate problem in Yes/No tests can be counterbalanced by including pseudowords with Greco-Latin roots and thus, students who
guess merely based on the familiar L1 root will be penalised. This, in his view, will underestimate passive recognition skills and provide a reasonably close approximation of active skills. The complexification of this, however, is that the test will perform unreliably across different cultures and in multiple-choice tests, such as the ones featured herein, the use of pseudowords is counterintuitive.

The fact that responses to cognates on Yes/No tests are more convoluted to interpret than a simple over- or underestimation of students’ vocabulary scores is best illustrated in Eyckmans (2004). She found that with some of the cognates, students demonstrated an erratic response behaviour, as they rejected them during the test, but provided a correct L1 translation in the post-test. One possible explanation indicates students’ scepticism towards these words as a foreign word. Other cognates were rightfully recognised, leading to the conclusion that cognates might not be simply and automatically accepted as real foreign words, despite the fact that it can only be assumed that they are part of their L1 vocabulary (Eyckmans, 2004). Furthermore, the exclusion of cognates from the results yielded a similar result for the items that students rejected on the Yes/No test, albeit, the number of correct translations declined. There was also an increase in the number of incorrect translations. Eyckmans (2004) interpreted this as evidence for the fact that cognates in the Yes/No test are not liable for a potential overestimation, the cognate effect can work in either direction, and the exclusion of cognates is not a solution for the problem.

In a more controlled study that utilised the Yes/No format, Petrescu and Helms-Park (2014 in Helms-Park and Drojnic, 2016) were interested in how Romanian and Vietnamese students recognise different types of cognates. As previously mentioned, cognateness can be seen as a continuum and thus they tested Romanian-English identical and similar cognates in form with semantical congruency (40 items), quasi-cognates (partial semantic overlap; 17), deceptive cognates (false friends; 21), pseudo-cognates (non-word that has a cognate in the L1; 20), non-cognate real words (20), and pseudowords (20) with no resemblance to
either English or Romanian. The class-by-class analysis showed that the Romanian’s advantage due to the cognates only became relevant in the case of semantically congruent items with a formal overlap. For all other categories of cognates and distractors, the advantage of cognates diminished. They infer that formally and semantically congruent cognates deserve attention as these stimuli are plausible to activate ‘semantic and orthographic/phonological aspects of the cognate in both of the languages’ (2016: 15).

As has been pointed out earlier regarding vocabulary test items, cognateness is not the only factor that can influence word difficulty. Most L2 vocabulary research has established a strong relationship between word frequency and accuracy (e.g. Milton, 2009; Schmitt, 2010; 2014). Therefore, one cannot assume that all cognates are recognised equally or at least with similar ease to non-cognates in their respective frequency bands. The reason for this is that cognates in the L1 or L2 of participants might occur more frequently than in the TL. The number of studies that take into account cognates and their frequency in relation to word difficulty is limited (e.g. Bennett and Stoeckel, 2014; Elgort, 2013; Stoeckel & Bennett, 2013; Stubbe, 2010). Bennett and Stoeckel (2014) were interested in the possible cause of differential item functioning in L2 English vocabulary tests based on loanword frequency in Japanese. To establish the frequency of the English words, the Corpus of Contemporary American English (COCA) was utilised and for Japanese loanwords, the Japanese subsection of Sketch Engine (jpTenTen11). Japanese loanword frequency emerged as a better predictor of word facility than the English one.

Since most of the studies conducted within L2 vocabulary acquisition address English, it is not surprising that a number of background language-specific tests, called bilingual VS tests, have been employed to further investigate cognates. The most notable is Elgort (2013), which aimed to compare the difference between the monolingual and Russian-English VST, and explore how response accuracy is affected by cognates. Relying on reliability evidence from Beglar (2010), Elgort split the VST in two 70-item tests, as this is less time-consuming,
and administered it to 121 Russian students of English. The ratiocination for using a bilingual version was that by replacing English definitions with L1 translations, some other aspects of L2 knowledge could be reduced (e.g. eliminating the definitions that the learner knows mean something else or misunderstanding the definition due to insufficient knowledge of grammar and/or syntax), anxiety of participants can be minimised, the test becomes less demanding, and therefore, more appealing for participants to respond to low-frequency items.

Elgort corroborates that participants’ responses were more accurate on the bilingual test and on higher frequency items. Furthermore, responses were also more accurate to cognates, however, the effect of the bilingual test decreased in the case of learners with higher proficiency. Another important finding in this study is regarding frequency and cognateness. First, it is important to note that in terms of cognates, the two versions of the test behaved similarly and the accuracy of responses to cognate and non-cognate items decreased for less frequent vocabulary items. This indicates that learners at intermediate or advanced proficiency have a good grasp of high-frequency items. However, in the case of cognates, the L1 knowledge plays a role and on low-frequency bands the cognate facilitation effect becomes noticeable up until the threshold that the L1 allows. Hence, students could not answer the majority of items, not even cognates, on the lowest frequency bands of the VST (Elgort, 2013).

Ultimately, it can be extrapolated from the above studies that background languages have an effect on English vocabulary development and assessment, especially if the known languages share lexical similarities with English. The question, then, becomes: in relation to other word characteristics, how much effect does cognateness have? Using a stepwise multiple-regression analysis, Willis and Ohashi (2012) tested 69 Japanese students on the first half of the VST and controlled for cognateness, word length, and frequency to find the most parsimonious model. They found that cognates are easier to learn and retain than non-cognates, and cognates of similar frequency and length in phonemes have a higher item
facility ratio, which is in contrast with Milton and Daller (2007), who only found a frequency effect. The reason for the discrepancy between the two studies probably lies in the way cognates have been conceptualised and operationalised as a variable. While Milton and Daller (2007) defined cognateness as the proportion of letters in common in English and French items, Willis and Ohashi (2012) employed a binary categorisation of cognate versus non-cognate items. It can be surmised that in the former study this type of graded measure is more a reflection of some of the underlying differences between the French (which tends to use more extraneous letters) and English spelling systems, which learners might be less susceptible to than the form-meaning connection that is non-detectable by checklist tests. Previous studies have also corroborated that identical cognates sharing a meaning are more likely to be recognised than words only sharing some similarities with or without semantic congruency (e.g. Petrescu and Helms-Park, 2014).

Willis and Ohashi (2012) also affirm previous studies showing that the more frequent and shorter words are more easily recognised. According to them, cognateness accounted for 54% variation in the model, while frequency and word length in phonemes explained 34% and 11% respectively. In a recent replication of this study, Reynolds et al. (2015) tested 330 Taiwanese English majors on all levels of the VST and found that polysemy contributed the most to the model (60%), followed closely by frequency (59%) and POS (26%). The reason for the lack of a significant cognate facilitation effect is probably that the number of cognates in the study was only 2.4% out of the 140 items in comparison to the 29% out of 70 reported by Willis and Ohashi. The results from these three studies confirm that given the proportion and distribution of cognates in the test and how cognateness is operationalised as a variable, word difficulty can largely influence the results of the test, due to learners’ prior lexical knowledge. What is common in all three studies, however, is the effect of word frequency, regardless of the L1 or test-formats used.

In order to further investigate the effect of cognates in classroom settings and
determine what influences their recognition and students’ perception of lexical similarities between their languages, a number of studies focused on the effect of cognate instruction. In the following the focus will be on such studies, starting with the most relevant one from the perspective of the present thesis.

**2.5 The effect of explicit instruction (EI) on cognates**

The aforementioned studies revealed that cognates can have a facilitative effect on FL learning and development. However, due to their special status in the learners’ background languages, the similarity between the languages, word characteristics, and degrees of cognacy of the actual items, congruent items can be problematic for testing. Furthermore, it has also been showed that cognate recognition is not a straightforward process. Learners often fail to notice them and this process can be influenced by age, proficiency, the degree of correspondence between the previously known word and its TL equivalent, and perhaps, even LL experience, or learning strategies. To address this issue and explore how LA can be facilitated by enabling students to rely on cognates, a number of studies have introduced an intervention to raise learners’ cognate awareness. In this section, the nature of cognate instruction, the role of cognates in the classroom, and their effect on vocabulary testing are reviewed.

Studies investigating the influence of a Romance L2 and English L3 from a VS perspective are limited, however. Molnar (2010), investigated the L2-L3 relationship and the effect of EI on cognate recognition in the case of Hungarian-Romanian bilinguals. Due to the cultural context and the test employed, her study provides a background to the present research, not only on a rationale but methodological level as well.

The study involved 26 Hungarian NSs from Romania with Romanian L2 and English L3. The participants were split into a study group (SG) that received EI on cognates prior to the test and a control group (CG). The aim was to explore students’ VS as measured by Nation’s version of the VLT (1983) on cognate level and overall. Students’ Romanian
Chapter 2: Literature Review

Proficiency was rated by their teachers. This was done for two reasons. First, Romanian grades were surmised inappropriate for the purpose as these merely reflect students’ knowledge in Romanian literature and/or grammar. Therefore, these would not provide a realistic picture of VS (Molnar, 2010). On average participants in the SG knew more cognates and performed more consistently too ($M = 61.57$ and $51.75$; $SD = 14.56$ and $18.65$). These differences were not statistically significant, however. With regard to the relationship between Romanian and English proficiency, the findings suggest that there is a significant moderate correlation between the scores overall (Pearson’s $r = .42$, $p < .001$) and the scores on cognates only ($r = .53$). As a post-test exercise, information on the students’ linguistic background was also collected.

Despite Molnar’s interpretation, the results indicate that based on learners’ age and proficiency, and perhaps the test used, the EI on cognates had no effect. One reason for this could lie in Hall’s (2002) postulation that since SLA is largely form-meaning driven, cognate recognition might be intuitive. Other evidence shows that young learners recognise cognates only after EI, in spite of the close L1-L2 lexical proximity (García, 1991). The ratiocination behind this is probably grounded in the findings in Nagy et al. (1993), which revealed a variation between cognate recognition scores in a reading task conducted with different age groups. They speculated that cognate recognition might be influenced by age and language proficiency. In summary, Molnar’s treatment group showed no difference as learners are adults and, therefore, they might recognise cognates automatically without the need for EI.

Another reason for the lack of effect of the cognate awareness-raising exercise might be that the VLT, in my view, is not sensitive enough to measure the effect of such short EI. Conversely, for cognate recognition in English, knowledge of the word’s meaning in Romanian is a prerequisite. Since Romanian vocabulary knowledge was not empirically tested, it is unwarranted to speculate whether learners reached the necessary proficiency in their L2 to recognise cognates. It can be argued that, perhaps, teaching cognate vocabulary
explicitly at different frequencies might be able to inflate the scores. However, for this a more in-depth and longer exposure would be necessary. Such instruction, or teaching material, could be informed by knowing the exact proportion of cognates on the different frequency levels (see Chapter 6).

One additional finding in Molnar (2010) to be considered is students’ comments on the EI and whether they found it useful. Molnar reports that ‘ten out of fifteen students found the instruction of structural relationships between the L2 and L3 words to be helpful in learning new lexical items’ (2010: 346). Furthermore, the majority of those students who responded negatively to the question of whether this method should be applied in the classrooms suggested their Romanian knowledge was not proficient enough to fully benefit from such interventions.

The other relevant aspect of this study that needs to be addressed is the L2-L3 link. The correlation between the scores overall and on the level of cognates suggests that students with a high L2 proficiency are prone to achieve higher scores in their L3. The association is even stronger on the level of cognates. This supports previous research emphasising that all known languages will interact with the TL and this interaction is most noticeable in vocabulary (De Angelis, 2007). Due to the limited number of participants and the subjectivity of teachers’ ratings, the results of Molnar (2010) should be treated with caution. In conclusion, the paper raised several questions that deserve attention. It is felt, however, that in order to draw valid conclusions, both languages should be tested empirically using compatible tools (addressed in chapters 4 and 5).

It is generally accepted that vocabulary knowledge is a core element of reading and comprehension (e.g. Anderson and Freebody, 1981; Bax, 2013; García, 1991). There is compelling evidence suggesting that L1 literacy and skills in reading can be transferred to L2 reading (e.g. Grabe, 1988; Tregar & Wong, 1984). In order to investigate the nature of students’ ability to transfer vocabulary to reading, Nagy et al. (1993) used Yes/No tests to
establish Spanish-English bilinguals vocabulary knowledge in both languages, a multiple-choice test for target cognate words and, after explaining what cognates are, a cognate-circling task. Following the two vocabulary tests and prior to the multiple-choice test, students received four reading passages that contained the target cognates. About half of the participants claimed to be more proficient in Spanish than English, which was confirmed by both vocabulary tests.

The results indicate that even a simple instruction on inter-linguistically congruent items allowed students to identify cognates in the passages and make a distinction between real, false, and non-cognates, albeit not all cognates have been recognised. The authors suggest that this can be due to the fact that students might not be able to recognise all cognate relationships or a lack of knowledge of the actual cognate in either languages. Moreover, it emerges from the findings that vocabulary knowledge in Spanish can be transferred to reading in English, and students scored highest on the multiple-choice test when they were familiar with the Spanish word and recognised it as a cognate as well. This is evidence for proficiency in a typologically similar language in conjunction with metalinguistic skills, such as cognate recognition, can enhance reading and comprehension in the TL. The fact that only about half of the cognates were identified, demonstrates that cognate recognition can be improved.

In comparison to Nagy et al. (1993), Proctor and Mo (2009) decided to embed interactive support within the reading passages to highlight Spanish-English cognate relationships to 16 bilingual (i.e. Spanish speakers learning in English) and 14 English monolingual pupils in grade four in the US. Subsequently, participants filled out a cognate awareness test, composed of 50% low-frequency English cognates that are high-frequency Spanish words and 50% non-cognates (48 items in total). The results reveal that while there was a significant difference between the monolingual and the bilingual group on the reading comprehension test, the monolingual students’ advantage diminished on the cognate-
awareness test. Although the total scores did not differentiate between groups, the number of correct answers to cognates indicates that bilinguals could profit from cognates.

Additionally, non-cognate scores indicate that in conjunction with the reading comprehension test, monolinguals had an advantage overall. However, the cognate facilitation effect significantly reduced the gap between the two groups, which is in line with Molnar (2008). These results, thus, highlight again the crucial role cognates play in vocabulary tests and reading comprehension. What the authors fail to do, however, is actually explain to what extent the instruction on cognates facilitated recognition as no baseline was identified or a bilingual CG. This way, it can only be assumed that the cognate instruction was aiding the bilingual group and, perhaps, due to the group sizes, the implications have to be treated with caution too.

The fact that cognates are recognised in some cases and not in others can be further exacerbated by orthographical and phonological differences (Helms-Park and Dronjic, 2012), and with the former, this is most pertinent in languages that use non-identical scripts. Helms-Park and Perhan (2016) investigated this in the case of Ukrainian (Cyrillic) and English (Latin). Their objective was testing the effectiveness of an ‘Explicit Instruction with Elaborative Processing’ (EI) and a reading-for-comprehension intervention (RC), in comparison to a CG that received no intervention. While the first group received EI on Ukrainian-English lexical similarities, read one text with the target cognates during which they had to infer and guess their meaning, the RC group received no EI but read two additional texts containing non-target cognate items.

In order to ensure participants in the three groups had roughly equal lexical knowledge at the onset of the study, the authors administered Schmitt et al.’s VLT (2001), which revealed no significant differences, except on the academic level. To measure the actual gains, in both the pre-test and post-tests the adapted version of the Vocabulary Knowledge Scale (VKS; Paribakht and Wesche, 1997) was used. Adapted in the present case meant
eliminating the oft-mentioned pitfalls of the test (e.g. not clearly distinguished or subjective response options; see Burton, 2009; Read, 2000; Schmitt, 2010), so they decided on turning the scores into four clearly distinguishable nominal levels (scales) where the focus was on providing meanings (Helms-Park and Perhan, 2016).

The pre-test indicated no difference between the groups, thus confirming the VLT’s results and the fact that at the onset of the study, the groups were comparable. However, at the offset, both post-tests indicated that the EI group outperformed the other groups, whereas there was no difference between the RC group and the CG. Within-group measures indicated that all participants achieved significant gains between the pre- and post-tests in providing meaning for cognates, albeit, this was not the case for sentential deployment. This was only significant for the EI group. To summarise, this experiment confirms that a short exposure to cognates, even in languages with non-congruent scripts, promotes cognate retention in terms of semantics and usage.

Finally, even though the study has obvious limitations (group sizes, test effect, reliability of post-tests, and the fact that in the case of the RC group it is not entirely clear whether the vocabulary uptake was only due to the increased exposure to English in terms of reading or the effect of cognates), the study serves as further evidence that cognates might not be automatically recognised in all cases. Thereby, receiving EI on cognates can have substantial pedagogical implications even in more distant languages, such as Ukrainian and English.

One common and underlying limitation in the empirical studies reviewed until now, and some of the following ones, is that learners’ different levels of proficiency, age, cumulative LL experience, and other competencies make it difficult to clearly distinguish between the role of cognates in testing or language development. These can be ascribed to subject- or learner-related factors, which are further exacerbated by language-related factors, such as the degree of similarity between languages or, on a lexical level, the degree of
overlap between cognate forms in particular languages. Often, it is difficult to find learners with a ‘clean slate’ in the TL to capture what exactly happens at the initial stages of acquisition and which aspects are most influential for development. One promising approach to circumvent some of these issues is to turn to artificial or novel languages, which admittedly, is not usually done for time constraints and potential ‘usefulness’ for students.

Bartolotti and Marian (2017) probed two different competing models (the scaffolding and accumulating models of lexical acquisition) to determine how the L1 or L2 of English-German bilinguals interact during novel language (Corbertian) learning, following a single experimental intervention. The novel lexicon they decided to teach comprised items selected based on two wordlikeness metrics: neighbourhood size and orthotactic probability. While the former one entails the number of words that ‘differ from a target in the substitution, addition, or deletion of a single letter’ (2017: 113), orthotactic probabilities measure the frequency of co-occurrence of unique letters or letter sequences in a language (e.g. st as the fourth most frequent onset bigram in English has a high probability value). According to the authors, the higher the frequency of such common features in the background languages and the TL, the easier the memorisation process and retention or retrieval will be. In other words, they contend that the French word stylo (pen), pertaining to its sublexical effects (st as a common word-initial bigram, and /sti/ and /lo/ as frequent sound sequences) facilitate encoding and retrieval. Relying on these two metrics plus human ratings, they categorised 48 Colbertian items into four distinct groups: congruent (high wordlikeness rating) items in both English and German, or neither; and congruent in only English or German.

The study posited significant in demonstrating that novel words similar to at least one of the languages are learnt and recognised faster than incongruent (i.e. low-wordlikeness ratings) lexis. Simultaneously, combined L1-L2 overlap did not yield significant differences compared to single-language overlap. This confers that novel words are likely to form links, or anchor, in a single background language, regardless of this being the L1 or L2 if similarity
to at least one of them exists, which serves as evidence for the scaffolding model. By contrast, the additive or accumulating model postulates that the new item anchors to both languages of the bilinguals, if a lexical or sublexical similarity exists. However, they found only limited evidence for this. Moreover, bilinguals’ VS was slightly larger in English and this is also the reason why English wordlikeness contributed more to accuracy in both recognition and production tasks.

In sum, the Bartolotti and Marian (2017) study is yet another addition to the myriad of evidence that inter-linguistic similarities on the level of lexis can benefit learners and accelerate LL, at least in the initial stages of acquisition. Compared to cognate recognition though, these wordlikeness ratings, as the authors implicate, facilitate and accelerate lexical acquisition as the higher the items’ rating is, the easier is to create (memory) links to the known languages. By links or anchoring, in actuality, they mean associative or keyword learning strategy (Shapiro & Waters, 2005) (i.e. the target item *duch* meaning ‘eyeglasses’ was either connected to the English word *duck*, as in ‘a duck wearing glasses’ or to the German *das Buch*, as in ‘using glasses to read a book’ [p. 132]), which lends additional support of cognates being an essential bridge between languages even if there is no meaning overlap as other wordlikeness factors still aid the memorisation and processing of the target words.

Classroom evidence shows that not only using the L1 in the lessons, but raising cross-linguistic awareness are received positively both by learners and teachers. Horst, White and Bell (2010), for example, followed a French and an English teacher to observe the L1 use in the classroom and provided students with cognate and other grammatical tasks where they had to identify languages similarities. Although the effect of such intervention was not empirically tested, both students and teachers received these awareness-raising exercises positively and based on the comments that students provided, it becomes obvious that students’ perception of English-French similarities or differences increased.
Other studies also highlight the fact that less advanced learners require training to capitalise on cognates and other cross-language congruencies (e.g. Jessner, 1999; Otwinowska, 2016; Ringbom, 2007). Conversely, other sources indicate that beyond a certain threshold, multilinguals intuitively, or consciously, utilise all the resources that are available to them (e.g. Jessner, 2006; Ringbom, 2007). According to Jessner (2008), this is evidence for the fact that multilinguals should not be seen as speakers of multiple L1s. Additionally, as argued throughout, focusing only on the L1 influence, in the case of multilinguals, might have a detrimental effect on the overall picture of vocabulary acquisition. The reason for this is, as the empirical studies presented in this section strenuously argue, that not only the L1 but the Ln can also play a significant role in the acquisition, processing and use of further languages. This is emphasised by the processing model of Meißner (2004), which postulates that if one Romance language has been acquired by the learner, it enables him/her to develop receptive skills in other Latinate languages. This can be directly linked with the arguments put forward by Bartolotti and Marian (2017), among others, that conceive of CLI as a positive attribute for LL.

In more distant languages, such as Polish and English, Otwinowska (2016) reports on a wide range of quantitative and qualitative studies conducted by her and colleagues, which explored cognate awareness and the use of cognates as a learning and motivational strategy. Based on the idea of affordances, widely used in psychology and adopted to multilingualism by Singleton (e.g. Singleton and Aronin, 2007), her overarching hypothesis is that ‘language learners have certain potential affordances at their disposal and that learning languages successfully depends on the learner’s capacity to perceive and utilise the linguistic affordances embedded in the studying environment’ (Otwinowska, 2016: 144). She emphasises that although metalinguistic and metacognitive aspects of LL influence the perception and capitalisation of cognates, even in the case of typologically and psychotypologically distant languages, cross-lexical similarities may potentially influence
The first and second studies presented in Otwinowska (2016) dealt with learners’ perception and knowledge of cognates. The findings in both studies indicated that more proficient students and those who spoke multiple languages, and therefore, had a better understanding of CLIs (cumulative LL experience), could rely more on linguistic affordances and, thereby, demonstrated increased awareness and knowledge of cognates. In line with Dressler (2000), this was also true for inferring meaning from unknown cognates, learning cognates compared to non-cognates and false friends, and as also indicated by other studies, the higher the orthographic similarity, the more noticeable the cognate facilitation effects were (Petrescu and Helms-Park, 2014). What is of interest here, however, is the extent to which cognate awareness-raising trainings influence the development of English lexical knowledge.

Her longitudinal classroom experiment comprised 72 hours of exposure for ‘total’ (as opposed to false) beginner adult learners of English, divided into an experimental and a CG. According to the methodology, the experimental group ‘did additional 28 exercises and communicative tasks activating cognates which were carefully incorporated into the programme’ (Otwinowska, 2016: 181). Although the group sizes were very small (8 in total), the results indicated that the experimental group translated faster and more accurately than the CG. Based on the oral test, students in the SG produced longer utterances, dealt with the topics more easily, and overall, showed increased fluency. Although it is not clear whether learners performed better due to the 28 additional exercises and communicative tasks, or because their increased metalinguistic awareness helped them in recognising more cognates that were not covered in the coursebook, it is certainly encouraging to learn that based on such exercises, students tend to became more confident in their FL dexterity and overall, seem to be more fluent and accurate in their production.

Contrary to the assumption that advanced learners should already possess an increased
metalinguistic awareness, previous studies showed some discrepancies towards cognate recognition. To explore this, Otwinowska (2016) reports on a longitudinal study that focused on the effect of cognate awareness training in receptive and productive tasks. The study surmised that multilinguals deploy a wider-range of vocabulary learning strategies and they rely more on lexical similarities between languages than bilinguals. The aim of the study was to investigate whether bilinguals (SG), who received EI, match their multilingual peers in comparison to the CG, where cross-linguistic similarities and cognate awareness was not part of the teaching.

Due to EI on cognates, the SG showed no difference to multilinguals, however, both groups significantly differed from the CG. Although all groups solved the same exercises, the CG only noticed cognates, while the SG could profit from effective and explicit knowledge of cognates, which is also in line with findings by Falk et al. (2013). Again, it is important to emphasise that from a pedagogical point of view, the assumption that adult and more proficient learners can automatically rely on their metalinguistic knowledge, LL experience, and cognate knowledge, might be misleading. Students can benefit indeed from explicit teaching of similarities and cognate rules (i.e. systematic changes in terms of inflexions and derivations) and this will manifest in other language modalities as well.

Otwinowska (2016) also focused on young learners with lower metalinguistic abilities. The treatment helped learners recognise more cognates and both their attitude and increase in cognate awareness was positive. This led her to intimate that raising awareness of cross-linguistic similarities should start as early as possible for it increases metalinguistic awareness, creates a positive attitude towards LL, which increases learners’ motivation as well, and ultimately, has a facilitative effect on the development of vocabulary knowledge and use in both adult and young learners at various proficiency levels.

In conclusion to this section, it appears that even if the potential correspondences between languages are more limited than in two Romance or Germanic languages, the
explicit focus on cognates enables interlingual inferencing, increases vocabulary uptake and use, which subsequently results in better language performance. The myriad of evidence for this opens up the enticing possibility to expand assertions that L2 acquisition is facilitated by the L1 to other background languages as well. Researchers such as Ellis (2006) and Schmitt (2008), inter alia, contend that learning the L1 tends to sharpen learners sensitivity to certain features and regularities of FLs. For multilinguals this sharpening, or cumulative learning experience and (meta-)linguistic awareness, results in learners being more attuned to new vocabulary if lexical or sublexical (orthographical and/or phonological) similarities allow for this. Maybe a consequence of this is also the evidence for that VS, after increased exposure or a certain age, tends to be a better predictor of lexical acquisition than working memory or alternative factors (Long and Shaw, 2000). The lexicon, however, cannot only be described in terms of size and links between the known languages, therefore the focus is now turned to lexical studies featuring a temporal measure.

2.6 Lexical knowledge and speed of lexical access

*Lexical access* is broadly defined as ‘the mental process of retrieving information about words such as orthographic, phonological and semantic properties’ (Foucart and Frenck-Mestre, 2013: 394). Researchers in the area of psycholinguistics conceive of it as the crux of multilingual research for it can be used to successfully explore the intricacies of the lexicon, determine how and at what speed different words, based on their various properties, are activated, and reveal the nature of specific interferences between different languages (e.g. Aparicio and Lavaur, 2014, Dijkstra, 2005). In the case of VS tests, such as the VLT and VST, the property of interest is the retrieval of semantic information about the lexical item, or meaning recognition. Laufer and Nation (2001) directly associate the lexical-semantic link with fluency. In their view, the speed at which the form-meaning mapping is done is an indicator of lexical fluency, and therefore, besides VS, a core element of language production.
Lexical knowledge and speed of lexical access

The direction of lexical access, in the case of the VLT is problematic, however. Since three meanings are provided (or in the case of the computerised version only one meaning at a time) and six test items (three responses and three distractors, see Chapter 3 for examples), it can be argued that learners rely on the meaning first and try to match it with the correct item. By contrast, in the case of the VST, where one item is presented and four possible meanings, the direction of lexical access is more straightforward. Nevertheless, converging with recent research, the time elapsed between seeing the form and answering yes/no in lexical decision tasks (e.g. Pellicer-Sánchez and Schmitt, 2012) or matching the item with its correct meaning (see Laufer and Nation, 2001; Tanabe, 2016) will be referred to as speed of lexical access.

Lexical access, then, in the case of the VST entails the process of recognising the written word (based on the stimulus and the non-defining sentence) and retrieving its semantic information from the lexicon, if known. The evaluation of correct retrieval of the items’ semantic property is based on whether subjects can accurately select the correct meaning out of the four options. Thus, lexical access, as measured by the VST, becomes inherently longer than in binary lexical decision tasks and for this reason, and to distinguish between the two methods, the measured time within this thesis will be referred to as response time (henceforth RT) as opposed to reaction time in lexical decision tasks.

Empirical studies that systematically investigate lexical access on VS tests are rather limited (e.g. Harrington, 2006; Laufer and Nation, 2001; Miralpeix and Meara, 2014; Pellicer-Sánchez and Schmitt, 2012; Tanabe, 2016). These studies extrapolated, however, two important factors regarding English, one being the relationship between the speed of lexical access and VS (size effect) and the closely related frequency effect. According to the latter, low-frequency items will require longer RT for the increased difficulty of the word, whereas the size effect postulates that more proficient students will also demonstrate lower RTs.
A study by Harrington (2006), for example, found that RTs systematically decreased with increasing proficiency (size effect). Similarly, Pellicer-Sánchez and Schmitt (2012), also using the yes/no format, found that participants not only gave faster responses to higher frequency words (frequency effect) but also to words they knew better. Concurrently, employing the yes/no format, Miralpeix and Meara (2014) found that the group with larger vocabulary provided significantly faster responses (size effect). Their findings did not confer evidence for the frequency effect, however, which lead the authors to suggest that there might be a VS threshold beyond which accessibility becomes a determinant factor.

Analogously, Laufer and Nation (2001) measured lexical-semantic access and came to the conclusion that lexical access becomes prominent once a certain vocabulary level is reached. That is, in order to demonstrate a significant increase in fluency (lower RT) at the 3k level, participants needed a VS around 5k. Furthermore, their study also indicated that, for NNSs, low-frequency vocabulary has a longer RT. Tanabe (2016) confirmed this – a large VS enabled quicker lexical access to meanings than vocabulary deficit.

The importance of measuring VS with a temporal measure lies in the fact that a fast-accessible lexicon is essential for the development of abilities such as speaking fluency or reading performance (e.g. Tanabe, 2016). Concurrently, Laufer and Nation contend that lexical dexterity is not sufficient for everyday language use, it ‘must be fluently available’ to learners (2001: 8). In this thesis, the interest in VS and RT also has a multilingual perspective. In particular, since previous studies in the area of psycholinguistics have found that lexical access is faster in the dominant languages (e.g. Aparicio and Lavaur, 2014) and cognates elicit faster and more accurate RTs (e.g. Elgort, 2013; Kroll & Stewart, 1994; Lemhöfer et al., 2004), these issues will be explored in the second study in conjunction with the above-described frequency and size effects on lexical access.
2.7 Third Language Acquisition: proficiency in multiple languages

Lastly, since research in several areas of applied linguistics have reached the conclusion that L3 acquisition has different characteristics to SLA (e.g. Cenoz, 2003; De Angelis, 2007; Herdina and Jessner, 2002), and vocabulary studies have pointed to the fact that lexical competence is a good indicator of not just language ability or proficiency (e.g. Laufer, 1992; Milton, 2009; Nation 2001, 2013; Schmitt, 2010; Stæhr, 2008), but also academic success (e.g. Daller and Wang, 2016; Laufer and Goldstein, 2004, Milton and Treffers-Daller, 2013), here I focus on proficiency in multiple languages, concurrently highlighting important aspects that make TLA qualitatively different.

The studies undertaken by Cenoz (2003; 2013) and Jessner (2008) are excellent examples of the fact that the field of education and language research, in general, have long distanced themselves from viewing bi- or multilingualism as a handicap or having a negative effect on cognitive development, but quite the contrary. As previous sections have already suggested, plurilinguals can have a cognitive advantage compared to monolinguals (see Bialystok et al., 2004). Evidence for this advantage come from two distinctive fields, namely bilingualism and SLA. These two fields conjoin in multilingualism or TLA as both pedagogical and sociolinguistic aspects are taken into account (e.g. Cenoz, 2003, 2013; Jessner, 2008). The effect of bilingualism on L3 proficiency, cognitive language development, the interaction between the background languages in different cultural, socioeconomic, and educational contexts have been investigated from several perspectives, such as general proficiency or specific linguistic aspects in terms of transfer or processing, therefore these studies also employed different, often incompatible, methodologies. For reasons of space constrictions within this thesis, the present section will take a holistic approach and focus on general proficiency in multiple languages, which are underpinned by the theories developed by Cummins (e.g. 1979, 1980, or 1991).

Cummins’ Developmental Interdependence Hypothesis is based on a wide range of
studies in different cultural contexts, involving different age and proficiency groups. It postulates that cognitive and academic skills in the L1 will be a significant predictor of skills in the L2, and therefore, proficiency, depending on the level (Threshold Hypothesis), will be transferable between languages (hence also the name Common Underlying Proficiency Hypothesis) (see Cummins 1979; 1980, or 1991). The Threshold Hypothesis predicates that bilinguals’ proficiency has to reach a certain level in both languages for a positive cognitive effect. In other words, high L1 proficiency is a predefining factor of the proficiency potentially attained or attainable in L2. Thus, a highly developed L1 competence with the appropriate amount of exposure to the L2 (all other things being equal: motivation, language predisposition etc.) will result in high L2 competence. Conversely, low L1 proficiency will yield to similarly poor or neutral consequences in the L2. Furthermore, he argues proficiency in the L1 and L2 can be used to describe L2 acquisition more accurately, not just regarding bilinguals, but FL learners as well. In relation to this, it has been contended that this effect, or CLI, is even more significant in the case of languages that are more related to each other (e.g. Cenoz and Genesee, 1998; Cummins, 1991).

Later, Cenoz (2003) extended this hypothesis to TLA and corroborated that an advanced bilingual proficiency yields cognitive advantages, which can manifest also in the L3. That is to say, if bilinguals (or simply L2 speakers) can transfer skills from the L1 to the L2, then, these cognitive, metalinguistic, communicative skills, or conceptual knowledge, can also facilitate the acquisition of an Ln. Evidence for this can be found in several sources. For instance, in relation to English L3, Cenoz (1991) and Cenoz and Valencia (1994), examined Basque-Spanish bilingual and monolingual Spanish students in secondary education and found that bilingualism exerted a significant influence on L3 proficiency measures in differing language modalities, such as listening, writing, speaking, reading, grammar, and vocabulary. However, they also conclude that general intelligence and motivation explained more variance in the model than bilingualism. This was later confirmed
Third Language Acquisition: proficiency in multiple languages

by Lasagabaster (2000), in the same cultural context and Sanz (2000) in relation to Catalan-Spanish bilinguals and Spanish monolinguals. Muñoz (2000) compared learners at different proficiency levels and found a significant correlation between Catalan, Spanish, and English, and concluded that high proficiency, or advanced aptitude (see Muñoz, 2014) in L1 and L2 yields similar proficiency levels in English L3.

In an exploratory analysis, Cenoz and Gorter (2011) compare writing samples in English (L3), Basque, and Spanish (L1 or L2) under the Focus on Multilingualism approach, which considers that LA should be approached from a multilingual perspective, exploring the language repertoire as one interlinked competence, rather than treating learners as monolinguals in each language (see Cenoz, 2013; Gorter and Cenoz, 2016). They examined content, textual organisation, vocabulary, language use, and mechanics (i.e. technical aspects of writing: spelling, abbreviations etc.) and found significant correlations between all variables with the exception of organisation. No difference was extrapolated between correlations in different language combinations, meaning that the relationship amongst the variables was approximately equally significant. Cenoz and Gorter take this evidence and suggest that ‘multilingual speakers share some skills across their languages’ (2011: 362). They also corroborate that well-performing pupils in mechanics or vocabulary in their L1 or L2 also tend to achieve high scores in the L3. Furthermore, examples of lexical transfer are provided in all directions by students. However, they report that transfer from the L3 to the L1 or L2 was less frequent than vice versa, possibly because learners are the least proficient in English and it is also the only FL. Finally, learners clearly employ the same multilingual competence when they approach the tasks as some of the strategies can be noticed in all three writing tasks of the same individual (e.g. writing descriptively, focusing on colours).

Although this is not an exhaustive account of the phenomena, there are two important messages that need to be taken further. The first is that proficiency in the previously acquired languages is a good indicator of L3 acquisition or state. The second is that due to the number
of languages bi- or multilinguals speak, their LL trajectory can be both quantitatively and qualitatively different from L2 learners. The main reasons can be summarised as follows.

Learners who are already familiar with two languages possess a more developed metalinguistic awareness, based on previous LL experience and the familiarity with potentially two different linguistic (sub-)systems, and therefore, can regard LL in a more abstract way, compared to monolinguals. Related to this is the idea that multilinguals are able to use several languages separately and effectively (Cenoz and Geness, 1998). Furthermore, L2 speakers, besides the cognitive and metalinguistic advantages, have also potentially developed more effective learning strategies on their journey that enable them to acquire a TL with improved efficacy. This can be linked to Cook’s notion of multilingual competence (1992). He posits that due to the interaction of languages known by the multilingual, their competence will be unique compared to the sum of monolingual competencies. Concurrently, Kecskes asserts ‘monolingual and bilingual children do not differ in what they do with language, but how they do it’ (2010: 100). By extension, this can also be related to not just bilinguals, but plurilinguals as well.

Another reason that largely differentiates L2 learners from monolinguals is the issue of affordances that has been exemplified in depth through the issue of lexical affordances, or cognates, in the previous sections, but can be further extended to grammar, pragmatics, and so on. These affordances, or ‘free rides’ (Cenoz and Todeva, 2009: 278), attribute learners with the advantage that many languages share similarities, and thus typologically close languages have a facilitative effect on their LL (see Cenoz, 2003, 2013; De Angelis, 2007; Jessner 2008; 2014). Such aspects of multilingualism link directly to the introduction, where the focus on plurilingual education and capitalising on learners’ multilingual competence were highlighted as part of the current language policies promoted by the COE and educational practices that focus on a holistic approach to language teaching and assessment (see Gorter and Cenoz, 2016). These issues are important as a general aim of this
thesis is to ask whether (1) it is appropriate to make lexical proficiency comparisons between learners without taking into account prior lexical knowledge and, by extension, potential sources of transfer, such as cognates. In addition, linking back to the Interdependence Hypothesis, if the ML or vocabulary knowledge in different languages are seen as interdependent, rather than autonomous systems, (2) is there a sizable relationship between lexical proficiency in the L2 and L3?

The following chapter brings the focus groups’ socio-cultural and LL background into perspective, with a special focus on Romanian as an L2. Subsequently, the benefits and limitations of lexical tests will be discussed together with the development of the respective Romanian tests that are used as instruments in the present studies.
Chapter 3: Measuring Vocabulary Size in the L2

Chapter 2 posited that the language system of the multilingual speaker is a complex and dynamic system in which the interlanguages, or subsystems, complement and support each other. It is well-attested in the literature that a certain connection exists across these subsystems on multiple levels and considering the whole language repertoire that is available to learners facilitates the understanding of LA in general, and the interaction between the languages in particular. Moreover, it has been also highlighted, and empirically substantiated in a number of studies (e.g. De Angelis, 2007; Molnar, 2008), that due to this interconnection, multilingual learners experience a LL trajectory that is both qualitatively and quantitatively different from monolinguals’. On a lexical level, it follows, then, that in order to unravel how cross-linguistic similarities between certain languages influence VS in general and, consequently, affect testing, the background languages of multilinguals should be taken into account as well. This, in turn, will enable the development of more theoretically grounded lexical tests that do not disregard interlingual inferences, but control for these in terms of cognates, and ultimately, have the potential to inform and improve language didactics and learning.

It has also been pointed out that proficiency in the known languages can serve as an indicator of the attained or attainable level of TL mastery. Therefore, the interest herein can be ascribed to the relationship between an L2 that belongs to the Romance family, namely Romanian, and English L3. Through this, it investigates the influence of prior lexical knowledge, not just on the level of proficiency, but in terms of potential sources of transfer, such as cognates, and the extent to which these special items within the lexicon impact VS tests. This raises the purportedly promising, albeit as yet scarcely subjected to rigorous substantiation, question of how VS can be meaningfully measured and compared in two or more languages.
To this end, in order to explore the interconnection between two languages within the ML in terms of lexical storage, on the one hand, compatible testing tools are considered a prerequisite. Such tools in Romanian were non-existent at the onset of this research. Therefore, developing appropriate tools and understanding the rationale behind vocabulary tests, and how these can be adapted to multilingual contexts is focalised within this chapter. Subsequently, in conjunction with the attempt to measure lexical knowledge in Romanian L2, understanding the cultural context and the language constellation that is central to the current research also becomes essential.

Hence, the aim of this chapter is to elaborate on the conceptual and empirical considerations that have been taken a priori and a posteriori to conducting the empirical studies. It thus provides a detailed account of the preliminary research on the development of the Romanian Word List (RWL) that the vocabulary tests are based on and then, it presents the instruments employed duly acknowledging their advantages and limitations. To achieve this, first, it provides a brief account of the importance of and rationale behind lexical frequency information in the case of teaching and testing vocabulary knowledge, whilst critically reflecting on how vocabulary instruction and development has been conceptualised in the Romanian context. Second, it reviews the VLT and the VST, presents their Romanian counterparts, and the compilation criteria that have been followed. The RWL and the Romanian versions of the VLT and VST are novel and unique to this research, therefore, by making them available for future research, these are also subjected to scrutiny and critical evaluation.

3.1 Cultural and educational context

In order to best explore the relationship between a Romance L2 and English L3 from a VS perspective, it was considered that a DLC in which the L1 has limited effect, i.e. it is not typologically approximate to either Greco-Latinate or Anglo-Saxon languages, allows a meaningful and in-depth analysis of the L2-L3 interplay. Since Hungarian is close to neither
English nor Romanian, the target group selected for this research was the Hungarian minority in Transylvania, Romania.

Figure 3.1 Ethnic minorities in Transylvania (2011 census). Adapted from Wikimedia Commons (Adrein, 2016)

Figure 3.1 displays the western territories of Romania, including those settlements that have a large Hungarian population, marked in green. According to the National Census of 2011, the total Hungarian population in Romania (based on ethnicity and mother tongue) was 6.5% (1,227,623), making them the largest ethnic minority, followed by Roma (3.29%) and Ukrainians, Transylvanian Saxons, Russians, and Tatars among others (< 1%). The majority of Hungarians live in Transylvania, which can be regarded as an autochthonous Hungarian region, making up 18.9% of the population in this, today, administrative macro-region (Institutul Naţional de Statistică [National Institute of Statistics], 2011). Hungarians are unevenly distributed though across all areas. For instance, in counties, such as Harghita
and Covasna, they reach 85% and 74% of the population respectively, while in other locations they live in diaspora as visualised in Figure 3.1 (Institutul Naţional de Statistică, 2011). It follows that bilingualism, and Romanian L2 knowledge, is defined differently in these regions. In territories where the majority of the population is Hungarian, the acquisition of Romanian is largely restricted to the classroom, and learners from such areas are less proficient and confident in using Romanian, especially in rural areas. Nevertheless, even though Hungarian can be considered a heritage language for some and, maybe the only, (native-)language for others, on a large scale, it can be inferred that Transylvania is a multilingual macro-region. Especially if English being compulsory from third grade (around the age of 8), is taken into account. In some cases, however, pupils start learning it in kindergarten, chiefly in urbanised areas and more advantaged institutes. A second FL (majorly German or French) is introduced then in the sixth grade, although this is not compulsory for Hungarians (schools can take this decision and most prestigious Hungarian secondary schools introduce it at least in the ninth grade [Degi, 2012]).

In general Romanians ‘know’ at least three languages by the end of school when they take the national Baccalaureate exam (age of 18-19) (Degi, 2012). In practice, this means that English is predominantly taught in at least two hours a week, however, in language focused specialisations, so called ‘intensive’ classes, this can go up to six hours a week.

Despite the facilitative effects of previously known languages, many teachers tend to disapprove the use of other than the TL in the classroom, thus allowing students to engage with the L2 or L3 to a greater extent and enhance their LL overall (see Horst et al., 2010). This is the aspiration in Romania as well (see Iatcu, 2000; 2005). This deeply entrenched and enduring attitude towards minimising the use of background languages has been seriously questioned though by the plurilingual methods promoted by the COE and pedagogic translanguging practices increasingly advocated in both face-to-face and distance teaching environments (e.g. Adinolfi and Astruc, 2017).
In the Romanian context, however, since most subjects are also available in Hungarian to NSs, when learning English, many have to translate and make connections between English and their L1, rather than Romanian. This is due to the fact that many English teachers in Hungarian schools are Hungarian NSs themselves and using a common tongue allows them to ensure students’ understanding. Although some of Iatcu’s (2005) conclusions are mainly based on anecdotal evidence, her study shows that young learners who had Romanian teachers of English produced higher scores on vocabulary tests.

Regarding teaching Romanian to minorities, it has to be mentioned, that the official state language is Romanian, although ethnic minorities (especially Hungarians) have the option to learn through the medium of L1. Romanian Language and Literature (four to five hours a week), Romanian Geography and History (2 hours a week each), however, remain compulsory subjects and they have been taught and evaluated according to NS standards up until very recently and it is still so in secondary school (Degi, 2012). In practical terms this means that Romanian is not regarded as an L2 for Hungarian NSs and, therefore, their language competence and knowledge is measured on the same official state test (National Baccalaureate) as Romanian NSs’, which is regarded by many as discriminatory (Pentek, 2011). This is mainly because such exams pose exactly the same requirements for NSs and NNSs and thus, impose extra pressure on both students and teachers. This is so despite non-native Romanians, obviously lacking comparable linguistic knowledge to NSs, have to cover the same amount of material set out in the national curriculum, which is always reflected in their poor results (Pentek, 2011). These issues seem to place L2 Romanian teaching and testing per se in a ‘pre-scientific era’ described by Weir (2003), where Cambridge Proficiency Examinations prior to 1945 comprised ‘the traditional, essay-based, native-speaker language syllabus including an English literature paper, the same as that sat by native speakers’ (Weir, 2005: 5).
3.2 Teaching and learning vocabulary in the Romanian context

Despite the fact that there is no political willingness to alter, or indeed ameliorate, the way Romanian is examined and taught for minorities (Pentek, 2011), in the past decades, a major shift in paradigm in certain publications and political agendas has been in evidence. This paradigm shift is mainly characterised by the conceptualisation of LL needs and the realisation that NNSs should be taught differently to NSs.

Regarding teaching and learning vocabulary, Platon, Burlacu and Sonea (2011) recognise this paradigm shift and emphasise that a rich vocabulary is an essential parameter for decoding or transmitting messages, while it also contributes to lexical diversity. Similarly, considering the Romanian teaching methods for monolinguals, Axenti and Verșina acknowledge the fact that teaching vocabulary in a systematic way allows learners to master the basic characteristics of word knowledge and ‘derivation with suffixes and prefixes has to be a means for enriching vocabulary’ (2009: 89).

Sîrghie (2009) also emphasises that high-quality written and oral communication is characterised by mastering orthographic, orthoepic (i.e. correctness of diction), and punctuation rules, and through continuous vocabulary development. Dina recognises vocabulary development as a key factor that is essential to the progression between the key stages of LL and reports that once ‘the essential vocabulary’ is mastered by learners, communication exercises are employed in order to further improve their knowledge (2013: 1034). The progress of newly learnt words from receptive to productive vocabulary is seen as an essential step in a number of publications (e.g. Axenti and Verșina, 2009; Sîrghie, 2009). However, the means for selecting vocabulary for teaching, how ‘essential’ vocabulary is defined, or the methods for evaluating students’ progress remain unclear.

With regard to the first two issues (word selection and ‘essential vocabulary’), Bărlea and Cerkez (2005), among many, make a distinction between fundamental or basic vocabulary and the ‘rest’. This ‘essential’ or ‘basic’ Romanian lexicon consists of
approximately 1,500 words that are commonly used by ‘all language users’ (p. 54), including body parts, colours, basic human actions, domestic and wild animals etc. It is questionable, however, whether for a foreign speaker of Romanian on the verge of commencing his/her academic studies in Romania, words such as *wolf, cherry,* or *ring* can be considered the *sine qua non* of LL.

Applying a more systematic approach, Biriș, Burlacu and Șoșa (2011) compiled a learner dictionary. Surmised as the ‘minimum vocabulary for Romanian’, it ‘comprises 671 entries, 1,410 pairs of antonyms and analogies, as well as over 2500 synonyms’. They claim that ‘it represents an efficient means of lexical acquisition since the antonyms are approached in a direct relation with synonyms and polysemous words’ (cover page). In other words, the selection criteria for items included in this list are based on semantic characteristics (items must have either synonyms or antonyms), thus ignoring concrete or abstract lexical items that might not fit these categories, but could still be considered essential. Furthermore, there is compelling evidence showing that compared to the thematic clustering of words (*e.g.* *eat, chocolate, hot, sweet, cup*), semantic clustering (*e.g.* *blueberry, strawberry, raspberry,* etc.) does not actually facilitate vocabulary acquisition due to the fact that the more distinct the words are the easier is to learn them (*e.g.* Erten and Tekin, 2008; Tinkham, 1993). By way of explanation, introducing lexica that belong to the same topic, but are of different POS, facilitates learning as opposed to synonyms and antonyms, or hyperonyms.

Regarding the third issue (evaluation of progress), Norel and Pop (2005) state that children’s vocabulary develops considerably by the end of preschool. This idea springs from Golu, Zlate and Verza’s (1992) Psychology high-school manual, which asserts that at around the age of 10-11, monolinguals know approximately 5,000 words, most of which is part of their active vocabulary. It is hard to tell, however, how they define what a word is and how this has been measured.

It becomes obvious from the aforementioned that vocabulary and lexical development
English frequency lists and measuring vocabulary size

is recognised as a key concept in learning and teaching Romanian. Although it is beyond the scope of this study to verify or judge the validity or efficiency of these claims, they certainly raise a number of questions: What is meant by ‘essential’ Romanian vocabulary and how useful vocabulary can be defined based on contemporary tools and knowledge? What is the target vocabulary for Romanian language learners? How can vocabulary knowledge (a ‘rich vocabulary’) be measured at different stages? With these questions in mind and prior to addressing them, the focus now turns to vocabulary research in the English language, as this is considered to provide the rationale for developing a Romanian frequency list.

3.3 English frequency lists and measuring vocabulary size

In order to place this preliminary research in the relevant academic context, this section addresses the importance of frequency information and lexical tests in the case of teaching, testing, and learning an L2. Since vocabulary is considered essential to LA, word lists and tests based on frequency information have been widely employed in mainly English contexts (e.g. Brezina and Gablasova, 2015; Meara and Milton, 2003; Nation, 2006; Milton and Treffers-Daller, 2013). The role of frequency information in lexical research is widely supported, insofar as, scholars working in the area state that it can be used to (1) set targets for students to acquire the essential coverage for understanding a wide variety of discourse and (2) quantify their VS. Together with the advancement of technology and corpus linguistics such empirical endeavours with the scope of understanding and improving FL education became increasingly viable and reportedly more accurate as well. However, despite the fact that vocabulary acquisition occupies a focal position in teaching Romanian, reliable Romanian frequency-based word lists and lexical tests are still missing.

What is exactly the rationale behind the importance of a lexical orientation? EFL studies corroborate the necessity of lexical development in and outside the classroom. The foundation for this probably lies in the oft-echoed argument by Wilkins: ‘Without grammar very little can be conveyed, without vocabulary nothing can be conveyed’ (1972: 111).
Researchers and language teachers also acknowledge that learners instinctively recognise the importance of lexical knowledge. For instance, Meara claims that ‘most learners identify the acquisition of vocabulary as their single greatest source of problems’ (1980: 221). Similarly, Schmitt (2010) argues that students, when abroad, refer to dictionaries rather than grammars. Therefore, finding the best methods to ensure learners make the most of their time and effort to increase their lexical knowledge is a challenging task (Milton and Fitzpatrick, 2014).

Vocabulary can be selected for classroom instruction in a number of ways and some of these are purportedly more successful than others. O’Laughlin (2012) and Schmitt and Schmitt (2014) found that many popular English course books only cover a limited number of useful lexical items and do not provide enough repetition of these. Although dictionaries are essential tools for LL and teaching, and their importance cannot be neglected, their success is dependent on the type of dictionary and the strategies learners employ. Using dictionaries to measure VS or select targeted lexical items for instruction have not been entirely successful, however. Nation and Coxhead assert that due to the dictionary spaced sampling method and the lack of a clear definition of the words used as standard counting units of measurement ‘almost all research on VS carried out during the twentieth century is virtually useless and at worst grossly misleading’ (2014: 338), which strengthens the argument for a more systematic and data-driven approach.

Since the appearance of various digital corpora, lexical frequency information has been found to provide practical resources for teaching and testing. Based on reliable quantitative methods drawn on corpus linguistics, most agree on that vocabulary can be divided into: high-frequency words (i.e. the most frequent 2000 words), general academic vocabulary (needed for academic studies), technical/specialised vocabulary (e.g. aviation or gastronomy), and low-frequency vocabulary (i.e. the ‘rest’) (e.g. Lessard-Clouston, 2012; Nation and Webb, 2011). Among the reasons for this lies the fact that a large number of
words is necessary for communication and language use, and frequency information can provide a realistic picture of lexis commonly employed in general language or even specific contexts. Moreover, as frequency lists and vocabulary tests are based on textual corpora, they represent authentic written and aural texts, making them even more efficient and attractive for both learners and teachers. The question that follows, then, is how can useful words be selected in a principled way from corpora or how can word frequency indicate lexical dexterity or deficiency?

The issue of lexical choice and vocabulary knowledge have a long and rich history in the area of (applied) linguistics. The first research dates back to a hundred years ago, when Thomson and Thompson (1915) utilised mathematical modelling to illustrate how English word repetitions in NS writing in terms of their frequencies and distributions have the potential to indicate individuals’ productive VS. The most influential discovery in this line of interest was coined by George K. Zipf in 1935, which is today referred to as Zipf’s law. This postulates: if all the words from a natural text or corpus are counted and ordered according to their frequency, the rank of each word will be roughly inversely proportional to its frequency of occurrence (e.g. Milton, 2009; Jarvis and Daller, 2013). In other words, in any given corpus there is an inverse relationship between the rank ordering of a word and its actual occurrence, and thus the word with rank one will be approximately twice as frequent as the second, three times more frequent than the third, and so forth. Zipf connected this with other power laws and argued that there is a natural tendency in language towards harmony and equilibrium, just as in music (1949). This, in practical terms, means that words are repeated naturally at certain intervals and, therefore, the word with rank one will occur at intervals of roughly ten words, the third one at every 30 (see Table 3.2) (Jarvis and Daller, 2013). Importantly, this repetitiveness is in a profound relationship with how humans formulate and perceive language as ‘word choice is carried out not only for the purposes of comprehensibility and information conveyance but also in order to hold the attention of the
listener or reader’ (2013: 3).

Consequently, learners engage with the TL in some way or form and in this process, they often meet highly frequent and some infrequent words, based on the aforementioned laws of language. This permits approximating for learners crucial vocabulary, the likeliness of encountering these, and most importantly, quantifying lexical knowledge and the necessary number of words for comprehension. Furthermore, since high-frequency words are encountered more often, the likelihood of mastering these words is greater as well (e.g. Milton, 2009). This is echoed in Gardner’s statement that ‘all vocabulary is not created equal’ (2013: 12). Subsequently, learners’ vocabulary knowledge preponderates in the high-frequency ranges, unless the TL or the input is highly specialised (e.g. aviation) (e.g. Milton, 2009; Nation, 1983; Schmitt, 2010). This also permits exploring, categorising, modelling, and overall exposing certain features of learners’ diction deployed in speaking and writing and the influences of background languages on this.

Since word frequency is quantifiable, there have been various attempts to estimate how many words are actually needed or known by users. Nation and Meara (2010) came to the conclusion that around 4-5,000 word families are required for an intermediate English proficiency and anything up to 9,000 for advanced. Attempts have been made to establish the VS of English NSs as well. Milton and Treffers-Daller (2013) review a number of studies in which educated NSs of English were reported to have VS estimations varying from 200,000 to as modest as 10,000 words. As mentioned earlier, the large range is due to different sampling methods. Their own investigation points to that even NS students’ VS may well be at the more modest end of the spectrum. Students’ knowledge of approximately 10,000 words (word families) at entry level and 11,000 in the final year shows a consistent variance around this figure. They also reason that despite students’ limited-range lexicon, the scores obtained tentatively explain the variation in their academic performance: ‘students with larger vocabularies tend to score higher in their assignments and exams and to obtain
higher degree classifications than those with smaller vocabularies’ (Milton and Treffers-Daller, 2013: 166).

In addition to VS, frequency data can be utilised to investigate the relationship between lexical knowledge and comprehension. The first 2,000 highly frequent English words, or the General Service List (GSL, West 1953), have been shown to provide 75-80% coverage of most texts. As Stoeckel and Bennett (2015) intimate, the most common words provide a disproportionately high percentage of coverage, resulting in valuable resources for prioritising over low-frequency words. Recast, learners, who attain the most frequent 2,000 words of English, will encounter around 20 unknown words in 100 in a general English text and will comprehend nearly 95% of spoken English (Adolphs and Schmitt, 2003). Nation (2006) investigated the requirements for the comprehension of novels and newspapers, and found that learners with a lexicon covering the 8-9,000 most frequent words in the British National Corpus (BNC) will have 98% lexical comprehension. Based on this and similar empirical evidence, Schmitt and Schmitt (2014) recommend that high-frequency vocabulary should be extended to the most frequent 3,000 words and below the 9,000 threshold the words should be categorised into a mid-frequency vocabulary. It is to be noted that these numbers mean word families and if these figures are translated into individual words, 8,000 word families actually consist of over 34,500 individual words (Nation, 2006).

Hitherto, the focal point of attention concerned the conceptualisation of frequency of word occurrence as a function of vocabulary knowledge, and how this has been operationalised and deployed in various contexts to establish lexical targets for students. Additionally, frequency lists can provide an essential resource for achieving the necessary vocabulary in form of graded readers or mastering academic vocabulary in EAP contexts. Word lists such as the GSL and the AWL (Coxhead, 2000) proved to be invaluable for this. The widespread use of these lists is also supported by the fact that Gardner and Davies (2013) recompiled the AWL (New Academic Vocabulary List) based on the COCA (Davies 2008).
Moreover, Brezina and Gablasova (2015), used four different language corpora to create a new GSL which features 2,122 core vocabulary items. Both the AWL and GSL were recompiled and refined by Browne and colleagues (see e.g. Browne, 2014).

Another notable example is the JACET 8,000 list (Aizawa, 2006). This, slightly larger list is compiled from a sizable learner-oriented material, which reportedly incorporates the majority of lexical items in other lists, more or less realistically distributing cognates and structure words, and it is considered more suitable for speakers of Latin-based languages (Miralpeix, 2008). Nation’s BNC frequency list (2006) has been used for creating several tests. Schmitt and Schmitt (2014) compared this to the COCA and found that the first 9,000 words provide coverage of just over 95% of this massive and diverse amount of data, which reiterates the importance of teaching high- and mid-frequency vocabulary.

These studies indicate that frequency lists and tests can provide a vast amount of information about the structure of a language, targets for learners, tools to reach these aims and quantitative, standardised tests for teachers and researchers to evaluate learners’ actual VS or monitor their progress over a period of time. However, as Macoveiciuc and Kilgarriff (2010) observe, Romanian is lacking a publicly available, large balanced corpus that would enable teachers and other stakeholders to improve teaching Romanian. This is probably also the reason for the virtually non-existent Romanian graded readers (one example is Arefu, 2014). To address this paucity, the compilation of a RWL is considered a prerequisite for designing Romanian lexical tools and exploring the VS of Romanian speakers. The reason for this is that, on the one hand, this bolsters the possibility of designing Romanian vocabulary tests that are surmised as comparable tools to the English ones utilised in the studies referred to in this thesis and widely in the area of lexical studies. On the other hand, by presenting the methodology for developing a new word list, the creation of the Romanian vocabulary tests is also more transparent, available for critical review and, most importantly, accessible for future replications and applications. The following section, therefore, will
outline the methods and procedures adhered to whilst compiling the Romanian Word List, and provide a detailed description of the list itself.

3.4 The development of the Romanian Word List (RWL)

According to Nation and Coxhead (2014) to eliminate the difficulties represented by the dictionary sampling method and develop a vocabulary test, a suitable frequency list is essential. This is ideally derived from a contemporary textual balanced corpus that represents real language from a wide range of subjects distributed proportionally, and from authentic written and oral sources. Besides English (BNC, COCA etc.) and a handful of other languages (e.g. French), such corpus is hard to come by, especially in Romanian.

One notable example is the 50 million word RoWAC (Romanian Web-as-Corpus) compiled by Macoveiciuc and Kilgarriff (2010) using the Web-as-Corpus method that can be accessed through the Sketch Engine (Kilgarriff et al. 2004). This was surmised unsuitable for the present purposes as the sources mainly represent journalism, the corpus cannot be regarded as a balanced corpus and, in relation to web-based resources, the texts’ authenticity is also questionable. Other Romanian corpora are either restricted for the public or to journalism, or represent lexicographic corpora that cannot be utilised for gathering and establishing frequency information. This research, therefore, employs the Romanian Balanced Annotated Corpus (ROMBAC; Ion, et al., 2012). However, contrary to its name, it does not include oral texts, and the written texts are drawn from largely formal scholarly areas. Nevertheless, it is still the largest Romanian corpus (not web-resource based) available to date. According to the authors,

discounting punctuation marks, [the ROMBAC] contains about 36,000,000 words evenly distributed into five genres: journalistic (news and editorials), pharmaceutical and medical short texts, legalese, biographies of the major Romanian writers and critical reviews of their works, and fiction (both original and translated novels and poetry)’ (Ion et al., 2012: 339).

The ROMBAC is restricted from the general public, but since it forms part of the META-NET (http://www.meta-net.eu), which is a network of repositories of language data and tools, and aggregates a wide range of resources for language research, it was possible to
gain access for use. The corpus is annotated and comprises XML formatted files, encoded in XCES schema. In order to extract the list, the trial version of Stylus Studio (http://www.stylusstudio.com/) was used and then to deal with large text files, Notepad++ (https://notepad-plus-plus.org/). The AntConc software was employed to compare word lists and determine coverage (http://www.laurenceanthony.net/).

A large cohort of researchers, such as Milton (2009) and Brezina and Gablasova (2015), and most recently Kremmel (2016), argue that lemmas provide a more reliable and realistic unit of word count than word families. The underlying reason behind this is that contrary to past assumptions and conventions, there is some evidence that not all members of a word family are known to the same extent by ESL learners or even NSs (e.g. Schmitt and Zimmerman, 2002). According to Milton (2009), it is commonly accepted that employing lemmas as the basic counting unit is most practical and reliable as it draws on the fact that learners will master frequent derivations and inflexions over the irregular or infrequent forms. Therefore, relying on lemmas as the unit of word count lends some confidence to the researcher that the tested item and its members indeed form part of the learners’ lexicon (see also Milton, 2009; Nation, 2006; Schmitt, 2010). Treffers-Daller, Parslow, and Williams (2016) have already put forward some evidence for using lemmas over word families in lexical diversity research. Further evidence and validation are still missing, however (e.g. Kremmel, 2016).

The previous section has also suggested that in the case of teaching and learning Romanian, a considerable emphasis is put on derivational and conjugational skills, indicating that due to the irregularities of the FL, not all members of a word family automatically form part of the learners’ lexicon and they might struggle with such skills. Consequently, it has been decided that at the initial stages of this research, the unit of measurement in the Romanian list will be lemmas instead of word families (see section 2.2.1).

The RWL developed from the ROMBAC corpus in two stages. At both stages the
following rigorous adjustment criteria were followed: punctuation marks, foreign words, numbers (including dates), proper nouns, abbreviations, duplications, and special and erroneous characters (e.g. %, ^, *) were removed. The reason for this was that once the script ran and it extracted the lemmas with their respective number of occurrence, the list contained all the above-mentioned characters. To attain a list for test-item sampling, such characters were deemed unnecessary. The raw data was cleaned up using Notepad++. Following this, the items on the list up to 14,000 had been checked individually to ensure there were no oversights.

The first stage of developing the list involved using the lemmatised list that is provided with the corpus. However, once the raw list which contained almost 500k items had been adjusted using the above criteria, even amongst the most frequent lemmas it was possible to encounter words from biology or medicine, such as glicemie (glycemia) or ribavirină (ribavirin). The word pacient (patient), for example, occurred over 36,000 times (frequency index) and thus it was the 27th (rank) most frequent word of Romanian. Due to the number of these words and the lack of possibility to filter them, these were left in the final list at stage one. However, these have been omitted during the word selection procedure for the Romanian version of the VLT, employed in the first study.

Stage two of the process was concerned with developing a more refined frequency list to be used for compiling the Romanian VST. This required the five different sections of the ROMBAC to be revisited separately.

Table 3.1 shows the distribution of tokens (running words), in the present case lemmas, and types over the five different sections. The total number of lemmas extracted from the ROMBAC is over 25 million tokens which amounts to just under 1 million types. It should be noted that the medical section is at least 26% of the corpus, and this explains the number of high-frequency medical terms in the initial list. As this section (3) is considered highly specialised and it is unlikely that a learner will account medical terms to this extent, a
decision was made to exclude this section to allow for more practical and general words in the list.

Table 3.1 The distribution of tokens and types in sections

<table>
<thead>
<tr>
<th>Section</th>
<th>Tokens</th>
<th>Types</th>
<th>Percentage in ROMBAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Journalism</td>
<td>1,922,109</td>
<td>50,945</td>
<td>7%</td>
</tr>
<tr>
<td>2 Literature</td>
<td>6,950,371</td>
<td>105,346</td>
<td>27%</td>
</tr>
<tr>
<td>3 Medical</td>
<td>6,783,005</td>
<td>362,782</td>
<td>26%</td>
</tr>
<tr>
<td>4 Legalese</td>
<td>6,269,543</td>
<td>248,354</td>
<td>24%</td>
</tr>
<tr>
<td>5 Biographies</td>
<td>3,716,031</td>
<td>223,592</td>
<td>14%</td>
</tr>
<tr>
<td>Total</td>
<td>25,641,059</td>
<td>991,019</td>
<td>100%</td>
</tr>
</tbody>
</table>

It also seemed logical at the design stage to eliminate compound words (*floarea-soarelui* – sunflower) and frequent collocates (*dinainte de* – before something; *conform cu* – according to) from the frequency list. This is justified by the fact that frequent collocates are mainly composed of two highly frequent words or at least one (e.g. *de, pe, la, cu, din* – *of, on, to, with, from* respectively), thus their presence would be duplicated in the list. Furthermore, the meaning of collocates can often, but not always, be deduced from the individual words that make them up.

Additionally, due to the number of variations, regularities, and frequency it was decided that nationalities (and languages) would be eliminated, thus allowing for a large number of more essential content words in the list. The ratiocination behind this is that due to the nature of the corpus, it heavily relies on literature and journalism, therefore such words were considered to occur in disproportionately large numbers with very low-frequency ranks and, similarly to proper nouns, at the item-sampling stage, they would be omitted anyway (i.e. neither the VLT nor the VST comprises such test items). Furthermore, as has already been articulated in the aforementioned, the higher the number of occurrences, the higher the chances that learners will master these items. By contrast, it can be argued that the amount exposure will evidently differ from individual to individual, due to the Romanian school curricula, in which a considerable emphasis falls on literature, geography, and history, words such as nationalities and languages potentially already form part of the target groups’
lexicon, and therefore, from a testing point of view, such items would unrealistically increase the item facility indices.

In addition, unlike English, Romanian corpora contain two different spelling systems, the old one (pre-1993) and the new one (post-1993). For instance, the words întâia (first) and întîmpla (happen) today are spelt as întâia and întâmpla (the ‘i’ with a circumflex accent has been changed to ‘a’ with a circumflex accent, except in the cases when the sound appears in a word-initial position, and the word sînt [1st person singular for to be] has been changed to sunt). From a methodological perspective, the decision was that words using the old spelling would be removed from the final RWL. The reason for this was the lack of a systematic way of altering the spelling to add the frequency count up for such words. Simultaneously, once learners mastered either version of these words, by pronouncing (or reading) them, they realise instantaneously what the words are (the difference is only orthographic). Therefore, either counting and testing, or teaching these words separately would be counterproductive. However, some of these are quite high-frequency words and thus, the coverage that the list provides can be considered an underestimate. For the current version of the list this issue is believed to have a somewhat minor impact and possibly in future versions it should be addressed accordingly.

The final RWL contains the 14,000 most frequent lemmas of Romanian. These single word items have been grouped into 14 different bands according to their rank. Thus, the first 1k band represents the most frequent 1,000 Romanian words and so forth. Admittedly, this clustering adheres to conventions rather than fully validated arguments (Schmitt, personal communication). Recently, Kremmel (2016) challenged this mostly pragmatic decision and suggested that based on the frequency distribution of lemmas (and also word families) a pedagogically reconsidered clustering might be more meaningful for lexical assessments. Relying on the abovementioned tripartite notion of vocabulary defined by Schmitt and Schmitt (2014), he suggests that high-frequency words provide more coverage, and
therefore, represent more essential items than lexis from subsequent bands. Thus, clustering high-frequency words into 500-word bands, followed by 1k-word bands for mid-frequency words and regarding as one band or completely omitting subsequent bands would warrant consideration as a more reliable sampling-method. Although this reconceptualization of frequency bands appears to be convincing, it is yet to be subjected to empirical substantiation. Furthermore, the English tests employed here are still based on 1k-bands, and thus, in order to allow for comparison, the proposed RWL also maintained the 1k-clusters.

It is common practice to validate similar word lists by comparison to different reference lists, and explore the coverage they provide (Nation, 2006). However, as the RWL is first of its kind, no reference or norm lists exist in Romanian to allow for such validation. Moreover, given the limited number of publicly available, balanced Romanian corpora, comparing the list to other corpora is also beyond the scope of this study. Nevertheless, a certain number of its features shall be explored.

First, it is important to discuss the frequency distribution of words. Numerous studies have shown that in any language high-frequency lexis represents a disproportionately large percentage of discourse (e.g. Nation, 2006). This is because English words such as the, of, get, give or Romanian words, such as de, și, eu, vrea (from, and, I, want respectively) are essential function or content words necessary to formulate meaningful sentences. Thus, their occurrence will be high in virtually any diction or corpus.

Figure 3.2 demonstrates the frequency distribution of the first 5,000 words. It appears that approximately the first 200 words occur in disproportionally high numbers in contrast to the rest of the words in the list. At around the first 500 words the frequency index stabilises and then gradually decreases. This tendency and the exponential distribution are similar in all languages due to the power law defined by Zipf, described above.
The development of the Romanian Word List (RWL)

Figure 3.2 RWL frequency distribution

This also facilitates a comparison between the BNC and the RWL (Table 3.2). The most frequent words (rank 1) in English and Romanian are *the* and *de* (*of*) and rank 2 words are *be* and *și* (*and*) respectively. Their individual frequency index is necessarily dependent on the size of the corpus that has been used. It is interesting to note, however, that there is a considerable drop between ranks 1 and 2, in the order of approximately 2 million words in English and 300k in Romanian. This difference becomes more moderate in the case of words with ranks 17 and 18 in comparison to English. Conversely, if words at around the 5k rank are examined, it is visible that the difference between their frequencies is minimised.

Table 3.2 Frequency comparison between the BNC and RWL

<table>
<thead>
<tr>
<th>Rank</th>
<th>Occurrence</th>
<th>Word</th>
<th>Rank</th>
<th>Occurrence</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BNC</td>
<td>RWL</td>
<td></td>
<td>BNC</td>
<td>RWL</td>
</tr>
<tr>
<td>1</td>
<td>6,187,267</td>
<td>826,777</td>
<td>the</td>
<td>5001</td>
<td>1,188</td>
</tr>
<tr>
<td>2</td>
<td>4,239,632</td>
<td>572,717</td>
<td>be</td>
<td>5002</td>
<td>1,188</td>
</tr>
<tr>
<td>17</td>
<td>675,027</td>
<td>138,934</td>
<td>with</td>
<td>5017</td>
<td>1,181</td>
</tr>
<tr>
<td>18</td>
<td>559,596</td>
<td>134,545</td>
<td>do</td>
<td>5018</td>
<td>1,180</td>
</tr>
</tbody>
</table>

In line with Nation (2006), the words’ frequency in the RWL decreases in a similar tendency to the BNC, which serves as evidence that the RWL is properly ordered. To further investigate this, it seems logical that the RWL is compared to subsections within ROMBAC.
What is expected is that high-frequency words should provide more coverage in another list or text than words at lower frequency levels.

First, however, the original list (stage one) used to construct the RomVLT contained a large number of medical terms, therefore, in order to verify that the two lists share similarities and the word selection criteria used for the RomVLT is reliable, it appears reasonable to compare the two lists.

Table 3.3 Comparison between the two versions of the RWL

<table>
<thead>
<tr>
<th>Level</th>
<th>Tokens</th>
<th>Token %</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,783</td>
<td>16.65</td>
<td>16.65</td>
</tr>
<tr>
<td>2</td>
<td>1,048</td>
<td>9.79</td>
<td>26.44</td>
</tr>
<tr>
<td>3</td>
<td>985</td>
<td>9.2</td>
<td>35.64</td>
</tr>
<tr>
<td>4</td>
<td>962</td>
<td>8.98</td>
<td>44.62</td>
</tr>
<tr>
<td>5</td>
<td>900</td>
<td>8.4</td>
<td>53.02</td>
</tr>
<tr>
<td>6</td>
<td>817</td>
<td>7.63</td>
<td>60.65</td>
</tr>
<tr>
<td>7</td>
<td>699</td>
<td>6.53</td>
<td>67.18</td>
</tr>
<tr>
<td>8</td>
<td>592</td>
<td>5.53</td>
<td>72.71</td>
</tr>
<tr>
<td>9</td>
<td>427</td>
<td>3.99</td>
<td>76.7</td>
</tr>
<tr>
<td>10</td>
<td>321</td>
<td>3</td>
<td>79.7</td>
</tr>
<tr>
<td>Not in the lists</td>
<td>2,176</td>
<td>20.32</td>
<td>100</td>
</tr>
</tbody>
</table>

Since the VLT only samples items up to and including the 10k band, Table 3.3 only compares these levels. Congruent with the expectations, as both lists are extracted from the ROMBAC (except that from the final RWL the medical section has been eliminated) the two lists do indeed share similarities. Overall, the frequency bands in the final list provide almost 80% of coverage. Furthermore, if the individual frequency levels are taken into account, it is noticeable that from the 10th level the percentages proliferate up to 16.65 at level 1. This proves that at least in relation to the original version, the RWL is properly distributed.

For exploring the coverage in literature, the RWL bands have been run through this section of the ROMBAC. Table 3.4 below indicates that the first 2k most frequent words in Romanian provide over 78% coverage of almost seven million running words and if band 3 is added, the coverage reaches 81%. In other words, mastering the first 2-3,000 words of Romanian would enable learners to demonstrate comprehension of at least four words in every five running words.
This is in line with the English language in which the most frequent 2k words provide around 80% coverage in a variety of contexts (see Milton, 2009; Nation, 2006; Schmitt and Schmitt, 2014).

### Table 3.4 The coverage of the RWL in the Literature section (2)

<table>
<thead>
<tr>
<th>Level</th>
<th>Token</th>
<th>Token %</th>
<th>Cumulative Token %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4,953,932</td>
<td>71.33</td>
<td>71.33</td>
</tr>
<tr>
<td>2</td>
<td>481,144</td>
<td>6.93</td>
<td>78.26</td>
</tr>
<tr>
<td>3</td>
<td>255,419</td>
<td>3.68</td>
<td>81.94</td>
</tr>
<tr>
<td>4</td>
<td>163,171</td>
<td>2.35</td>
<td>84.29</td>
</tr>
<tr>
<td>5</td>
<td>119,148</td>
<td>1.72</td>
<td>86.01</td>
</tr>
<tr>
<td>6</td>
<td>96,119</td>
<td>1.38</td>
<td>87.39</td>
</tr>
<tr>
<td>7</td>
<td>73,364</td>
<td>1.06</td>
<td>88.45</td>
</tr>
<tr>
<td>8</td>
<td>57,972</td>
<td>0.83</td>
<td>89.28</td>
</tr>
<tr>
<td>9</td>
<td>46,809</td>
<td>0.67</td>
<td>89.95</td>
</tr>
<tr>
<td>10</td>
<td>40,609</td>
<td>0.58</td>
<td>90.53</td>
</tr>
<tr>
<td>Not in the list</td>
<td>657,609</td>
<td>9.47</td>
<td>100</td>
</tr>
</tbody>
</table>

Additionally, if lower frequency bands are considered, it can be observed that the total amount of coverage the 10k most frequent words of Romanian would provide is just about 91%. Turned into comprehension figures, this still means that on average learners would encounter one unknown word in every ten running words. This could be surmised a manageable amount. However, learners at this level would still struggle with most ungraded literature texts if only these words were known. On the other hand, as interlanguage develops, learners will incrementally attain a large number of proper nouns, increase their knowledge of compound words and become more capable of recognising and producing collocates. Since these have been eliminated from the RWL and proper nouns are highly frequent especially in the area of literature, the total coverage that the RWL provides might be well over 91%.

Consequently, based on the frequency information extrapolated by the RWL, it is argued here, following Schmitt and Schmitt (2014), that up to and including the 3k most frequent Romanian words should be regarded as high-frequency vocabulary and up to and including the 10k level as the Romanian mid-frequency vocabulary. As the findings suggest, these lexical items provide essential coverage for language learners. Anything above these
levels should be regarded as low-frequency vocabulary.

According to Nation (2006), inter alia, even in widely researched languages, such as English, for which several large, publicly available balanced and learner corpora are concurrently available, developing such word lists is still problematic. Among the reasons for this is that the technical advancements available today still make it difficult to distinguish between homonyms. Thus, Romanian words such as *Capitan* (proper noun) and *capitan* (*captain*) are either considered as one word, which in the case of quotidian words can account for a sizeable difference, or as separate words, recognised by the capital letter, in which case sentence-initial common nouns will be counted as proper nouns.

Despite such issues, the above limitations and methodological flaws, the RWL is the first frequency list that can be construed as a potentially useful source for teachers, learners, researchers, and curriculum designers alike. It can serve as an indication of words that they should concentrate on and as a reference list for future studies in Romanian. Moreover, it appears that the list demonstrates similar tendencies to English frequency lists, which provides some confidence about its reliability regarding its frequency distribution, however, its representativeness of spoken or ‘real’ language remains questionable. Lastly, it is certainly expected that it proves to be robust in its practical application as well, i.e. in designing the Romanian versions of the VLT and VST.

Following the development of this prototypical tool, the focus now turns to the instruments used for the empirical studies featured herein. As acknowledging that the nature of the tests per se are crucial elements of lexical assessments, the second part of this chapter offers initially a critical review of the English tests, followed by the development of their respective Romanian versions based on the RWL. In turn, these sections purposefully move forward the detailed instrument sections of chapters 4 and 5 into a concise part of the thesis.
3.5 The Vocabulary Levels Test (VLT)

As explained, vocabulary testing provides researchers and teachers with reliable quantitative information to draw meaningful comparisons between the learners’ overall language knowledge, comprehension, and FL performance. It also permits effective individual and group comparisons between learners at different levels, even in the case of a large number of participants, provides means for deciding what type of words to teach and at what frequency levels, and aids efficient evaluation of progress (e.g. Milton, 2009; Nation, 2001; Schmitt, 2010; Schmitt, Schmitt and Clapham, 2001; Webb and Sasao, 2013). Furthermore, frequency-based lexical tests have proved to be practical and attractive tools for teachers as they can be corrected and analysed relatively quickly (e.g. Fitzpatrick, 2007; Fitzpatrick and Clenton, 2017). Nation’s VLT (1983) accomplishes these by various means.

Firstly, the VLT, as an English diagnostic tool, is constructed in such a way that it allows testing at five different frequency levels, namely the 2k level (the first 2,000 most frequent words in English), 3k level, 5k level, the University Word List and finally, the 10k level.

Secondly, each level contains six clusters and each cluster contains three test items and three distractors, which theoretically equals to one item (key or correct answer) and five distractors within one cluster for the first item, four distractors for the second item, and three distractors for the third item. In practice, as the example below shows, testees have to match the three definitions provided with three (keys) of six response items. This way, each level is in fact testing for 36 individual words, including nouns, verbs, and adjectives distributed proportionally. According to Nation (1983; 2001), this format allows for an easy design and marking, reduces the chances of guessing correctly to a minimum, facilitates the testing of a relatively large number of words in a short time, and enables students to rely on a wide range of knowledge to make their choices.
Chapter 3: Measuring Vocabulary Size in the L2

Extract from the VLT (Nation, 1983)

1 business
2 clock
3 horse
4 pencil
5 shoe
6 wall

part of a house
animal with four legs
something used for writing

The format also allows for flexibility in the sense that it does not have to be administered in its entirety. Teachers can decide which level they want to test, and interpret only those results. This is especially helpful with beginner students and where vocabulary uptake has to be optimised for academic purposes for instance.

The VLT and its revised versions have been used in a wide range of contexts (e.g. Akbarian, 2010; Staehr, 2008; Webb, 2007; Webb and Sasao, 2013) and a number of validation studies concluded that it is a reasonably good measurement of lexical knowledge (Beglar and Hunt, 1999; Read, 1988; and Schmitt et al., 2001). Kamimoto (2014) investigated, however, the local item independence and found, contrary to Schmitt et al. (2001), that the VLT’s format violates the assumption that an item within a cluster ought not to be dependent on a response to another item. It has to be acknowledged that in Kamimoto’s study three clusters are collapsed into one, thus increasing the number of definitions to nine and the number of items and distractors to 18. Although this format might reduce the amount of guessing, which is around 18%, as reported by him, it also increases the cognitive load that students require to fulfil the task, which, in return, could result in reduced scores.

Nevertheless, according to Meara (1996) and Schmitt (personal communication), the VLT is the closest there is to a standard test and the above reasons are just some of the many to support the fact why the VLT is so robust and why it has been utilised and adopted by several researchers, not only in English but incarnated in several different languages (see Webb and Sasao, 2013). Schmitt et al. (2001) revised Nation’s test and implemented two important changes. One is the increase in the number of tested items from 18 to 30 on each frequency level in attempt to escalate its reliability. The other one is the use of the AWL to
The Vocabulary Levels Test (VLT) test academic vocabulary (UWL in Nation’s version). Although this is considered an improved version due to more reliable scores, I employed the original one for two different reasons.

On the one hand, the original test was used by Molnar (2010), and consequently, to allow for replicability and comparison between the two studies, the same test had to be deployed. On the other hand, as the new version features an increased number of items, designing a similar test in Romanian and administering them both to the same group simultaneously would be overly time-consuming for students, thus they could potentially lose interest.

Apart from the benefits of each version of the VLT, their limitations have to be addressed as well. As with all vocabulary tests, they do not allow to draw far-reaching conclusions on overall proficiency as they feature only a limited number of items (e.g. Fitzpatrick and Meara, 2004; Schmitt, 2010). However, the VLT allows comparing individuals and groups on the same task (e.g. Moreno, 2010). Added the fact that since only certain levels are tested, the test cannot measure or estimate overall VS to a high degree of precision. Furthermore, only one meaning is tested, which is problematic, as Laufer and Goldstein (2004) also contend, because weak or partial knowledge of the tested item can result in a correct hit too. Moreover, as Web and Sasao (2013) assert, the test does not include the first thousand most frequent words of English and the frequency lists used for word selection seem to be outdated (e.g. GSL in Nation’s version). Although these points made by Web and Sasao are valid, for the study presented here tests participants who are English majors, it is assumed, that the 1k level has already been acquired and thus no testing is required on this level.

A further limitation regarding the scoring of multiple-choice tests, but not just, has to be mentioned. A common feature of vocabulary tests is that if the item is marked correctly, knowledge of that particular item is assumed and overall vocabulary knowledge is
approximated. However, word knowledge is more complex and multi-faceted. Kremmel and Schmitt (2016) investigated this aspect and came to the conclusion that it is perhaps not warranted to extrapolate further than the form-meaning knowledge of the actual item and the employability of these words productively, sentential deployment, or in comprehension tasks have to be assumed with caution. This is, in actuality, also one of the reasons why form-meaning tests were selected for the current research as, compared to yes/no tests, with multiple-choice formats there is at least some degree of confidence that learners are familiar with the meaning of that particular word. Although this cannot be taken entirely at face value either, leading to another limitation of such formats – guessability.

Evidence from some recent studies suggests that there is a 17% probability of subjects simply blind guessing the correct response on similar formats to the VLT (e.g. Kamimoto, 2008; Webb, 2008). Stewart and White (2011) employing a formula that relies on elementary probability theory, arrived at a similar conclusion, according to which, regardless of ability, the six-choice format leads to an average score inflation of 16%. ‘Regardless of ability’ here means that once students reach beyond a certain threshold, their guessing rate slightly diminishes. The following two studies conceivably circumvent the issue of guessing as the main aim is comparing lexical knowledge in the L2-L3, and exploring the impact of cross-linguistic similarities without drawing far-reaching conclusions in other language modalities, such as writing or speaking. Thus, despite the rate of guessing contended in the above studies, it is hypothesised that this will be similar across the two non-native languages, which, consequently, facilitates a close comparison between the two.

Moreover, as Molnar (2010) asserts, the original VLT contains a considerable number of Romanian cognates, 48 out of 90 tested words. Since the aim is to explore the effect of cognates on lexical tests, having a roughly 50-50% proportion allows for a differentiation between cognate and non-cognate item facility indices. It follows that although an up-to-date frequency list might be more appropriate, I considered that Nation’s (1983) VLT is most
suitable for the current purposes based on the above arguments.

Finally, the VLT as a typical form-meaning mapping test with a multiple-choice matching format, which employs a stratified sampling method from different frequency bands, together with its Romanian parallel version, allows the investigation of lexical storage and retrieval in the L2 and L3, and empirically exploring the effect of cognates on such lexical tests.

### 3.6 The Romanian Vocabulary Levels Test (RomVLT)

This sub-section presents the compilation criteria of the RomVLT (Appendix 2), discusses the criteria followed, highlights the challenges involved in designing it, and points out some important contributions it brings to the field.

Similarly to the VLT, the RomVLT is based on frequency information, which in this case is provided by the RWL. Since the items and clusters in the English test are sampled from the 2k, 3k, 5k, UWL, and 10k levels, the objective was to sample words for the RomVLT from these levels as well. Nation (1983) positions the UWL around the 6k level and, as there is no Romanian equivalent to this list (containing high-frequency academic words), when designing the RomVLT, the words for this level were sampled from the 6k level. Overall, the tests consist of 180 words, featuring five different levels, each including six word clusters and each cluster containing six words (36 in total on one level). An example is provided below.

**Extract from the RomVLT**

1. semnătură (signature)
2. supraviețuire (survival)
3. conversație (conversation)
4. dialog (dialog)
5. rinichi (kidney)
6. intensitate (intensity)
7. clădire (building)

Note: translations are only provided herein (see Appendix)

Throughout the item selection procedure, it has been ensured that the number of the
noun, verb, and adjective clusters match the proportion in the VLT. Specifically, the words from the stratified sample follow a 3:2:1 noun, verb, and adjective ratio respectively. Thus, on each level separately, the item clusters were selected in such a way that they reflect this ratio defined by the VLT. Additionally, for the aim of the test was to enable exploring the recognition of inter-lexically congruent items, the number of cognate words on the VLT were counted on each level (see Table 3.5 below), in both the list of tested words and the distractors. For example, each level contains 36 items (keys and distractors) and on the 2k level 15 are Romanian cognates and ten out of these are tested. The number of cognate words on the RomVLT were selected according to this. Table 3.5 reflects the distribution of cognates on each level.

Table 3.5 Distribution of cognates on the tests

<table>
<thead>
<tr>
<th>Level</th>
<th>Cognates out of 36 items</th>
<th>Cognates tested out of 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>2k</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>3k</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>5k</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>UWL/6k</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>10k</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>87</strong></td>
<td><strong>48</strong></td>
</tr>
</tbody>
</table>

Each level contains six clusters and each cluster contains 3 items and 3 distractors, there are 36 potential items on each level and 18 words that are test items (keys or correct responses)

The criteria followed for defining cognates was the one described in Chapter 2. According to this, items were deemed as cognates if their meanings are overlapping in English and Romanian, regardless of being loanwords or internationalisms, share a reasonable amount of formal similarity (i.e. only orthographic similarity and at least 60% of the letters are identical, disregarding diacritics or circumflexed characters). Furthermore, a small number of words, such as *sport* and *trumpet* are cognates in Hungarian as well. However, the number of these words is minimal and all are regarded as high-frequency words, so it is surmised that these will not have an impact on the overall scores, but will be returned to in the results section. The English test included no false cognates. Therefore, no
false cognates were included in the RomVLT either (see Chapter 6 for a discussion).

Reflecting the other criteria that have been considered, described in detail in Schmitt et al. (2001: 59), I ensured that:

- Keys and distractors were single words, instead of definitions
- To minimise reading time, reduce cognitive load and allow for efficient presentation, the definitions were kept as short as possible
- Besides a simply worded definition, the words within the definition are from higher frequency bands, thus a band 6 item on the RomVLT should only contain more frequent words in the definition (up to 6k)
- The items selected were base forms of the respective lemma
- Overall, words within a cluster were selected that they are different orthographically in form
- Definitions were formulated that they avoid use of the target item
- Items within each cluster were meant to be independent, i.e. different meanings

The definitions for the items and cognate status were checked mainly using the Romanian DEX (Dicționarul explicativ al limbii române – Explanatory [monolingual] Dictionary of the Romanian language), available online at https://www.dex.ro and major English dictionaries available online. Finally, the test has been piloted with two educated Romanian NSs in order to ensure that there are no errors and omissions, the items are independent, and the definitions are not misleading.

Should the results from the ensuing studies indicate that the obtained results are reliable and the test is capable of capturing the assumed theoretical profile of a Romanian learner, would further substantiate that the RWL is properly ordered. This, in turn, would lend some degree of certainty that both the RWL and the RomVLT could be useful in a wider application for assessing the lexical knowledge of ethnic minorities in Romanian at different ages and proficiency levels. It would allow the comparison of L2 and L3 knowledge at a lexical level, facilitate the design of longitudinal studies investigating lexical acquisition and so forth. The test results obtained are scrutinised in terms of reliability and validity in section 4.8.1.
3.7 The Vocabulary Size Test (VST) and developing its L1 and L2 versions

The studies presented in this thesis are incremental in nature with regard to their development. Thus, to further explore the RQs, investigate the results in Molnar (2010) and the first study (Chapter 4) in a more controlled, online (both in terms of real-time processing and web-based) experimental paradigm, it was decided that instead of the VLT, the second study (Chapter 5) will opt for the VST (Nation and Beglar, 2007). There are several reasons for this, some of which have been expanded on previously, and ensue in Chapter 5, but some additional information is afforded here. First, compared to the VLT, the format of the VST allows an online measure where the speed of response can be recorded. Although this would be possible with the VLT if only one item out of three was presented at one time (see Laufer and Nation, 2001), the fact that participants would have already seen the five distractors, during the second and third items, the RT measures would be convoluted to interpret (Harrington, 2006).

Second, as Nation (2001) also admits, the VLT is a diagnostic tool per se that allows teachers to investigate where students need to focus, and perhaps how many academic words they know. By contrast, the VST was developed as a comprehensive written VS test that measures vocabulary knowledge on all fourteen levels of the BNC list (Nation, 2006), testing ten items at each level. An example of the item format is provided below.

**Extract from Level 1 of the VST (Nation and Beglar, 2007)**

1. SEE: They saw it.
   a. cut        b. waited for
   c. looked at  d. started

The VST has also been employed widely in different studies in which its validity and reliability were explored (e.g. Beglar, 2010; Gyllstad, 2012; Gyllstad, Vilkaite and Schmitt, 2015). Compared to yes/no or checklist tests, which use a self-report binary-choice format, the VST, as a prototypical four-option form-meaning mapping multiple-choice test, allows
for testing a relatively large number of diversely sampled words by controlling for difficulty. According to Nation (2013) and Beglar (2010), this is achieved through selecting items that demand roughly the same degree of knowledge for each item. The consistent item-writing procedure of one set of item and four options ensures consistency and familiarity with the format across different cultures. Furthermore, the dichotomous marking criteria, where each item is marked either correct or incorrect, provides objectivity and reliability to the scoring, and lends a limited degree of certainty that the meaning of the item is known (Beglar, 2010: 26).

Studies undertaken by Wesche and Paribakht (1996) and Gyllstad et al. (2015) raise some concerns regarding multiple-choice tests. Among these are that test-takers could rely on their knowledge of the distractors, thus eliminating incorrect responses and increasing the chances of guessing. This, according to Stewart (2014), and confirmed by Gyllstad et al. (2015), could mean an inflation in scores as high as 25%. As with the VLT, it is hypothesised herein that students’ guessing rate or the overestimation of their vocabulary scores will be proportional to their proficiency on all three vocabulary tests. This means that although the VST is not unproblematic, the Romanian (ROVST) and Hungarian (HUVST) versions will have the same characteristics, therefore enabling a comparison of VSs and more importantly, an investigation of the relationship between these languages on a lexical level. Moreover, the lack of standardised vocabulary tests poses an unresolved conundrum and a challenge to researchers in the sense that they have to weigh the advantages and disadvantages of each test and find the best fit for their purposes (e.g. Schmitt, 2010).

Third, Beglar (2010) intimates that using only half of items (70), the scores still remain valid and reliable, which allowed to reduce the number of items in the tests. This was required as testing 140 items in three languages (420 items) would considerably reduce participants’ interest and motivation in taking part in the study. Conversely, based on Schmitt et al. (2001), this would be impossible with the VLT as they call for increasing the number
of items in the VLT to maintain reliability estimates. Additionally, at the design stage, I was
fully aware that, according to the Classical Test Theory, and also confirmed by Gyllstad et
al. (2015) and Schmitt et al. (2001), in order to increase the validity of the results ‘the more
items the better’ approach should be adopted. However, as the aforementioned emphasises,
the purpose of the current research is to explore the interconnections between multiple
languages and the intricacies of the multilingual ML. This way, retaining the entirety of the
VST was deemed impractical.

To this end, the VST was split into two parts following Elgort’s (2013) criteria, and
items that were classified unreliable or misfitting according to Beglar’s (2010) item analysis
were eliminated too. These problematic items and false cognates (three on the entire test)
were replaced with the same POS wherever possible. The thus remaining 70 items
(henceforth ENVST) distributed equally on the 14 frequency levels have been checked for
cognateness in Hungarian and Romanian. Cognateness in this case was further broken down
into bi- and multilateral cognates. Bilateral cognates were conceptualised as interlingual
homographs that share a similar meaning across two language combinations only (EN–RO
or HU–RO), whilst multilateral cognates were items that are present in all three languages.

Additionally, to facilitate a more in-depth analysis of the factors that can potentially
influence item difficulty, and consequently form-meaning recognition on the tests, after
Willis and Ohashi (2012) the following properties were also controlled for: word length,
operationalised as the number of syllables and POS in terms of nouns, verbs, and adjectives
(see Appendix 6 for a complete item property table). Given the format of the VST, the four
options and a non-defining sentence, these factors were considered to be the most influential
both in terms of the temporal measure (accessibility) and test scores (accuracy). Therefore
other item-characteristics that could potentially have an impact on lexical recognition, such
as imageability, concreteness, or animacy, for example, were not controlled for. This
methodological decision is premised on the consideration that in the case of lexical decision
tasks there is compelling evidence towards concreteness or animacy (e.g. De Groot and Keijzer, 2000), for instance, as having an influence on lexical access. However, the majority of time in searching for the correct answer from the four possible options, which are in some cases constructed of multiple words, and the non-defining sentence, is not likely to be determined by such word characteristics. Consequently, word frequency, cognateness of items, word length, and POS are surmised to directly affect word-familiarity, and therefore, might contribute to a greater extent to meaning recognition.

Strictly following Nation and Beglar’s (2007) criteria for the item, response and distractor selection, for the ROVST the RWL was utilised. For the HUVST Varadi’s frequency list was used, which is derived from the Hungarian National Corpus (2002). During the compilation stage, every effort has been made to select distractors and responses that come from the first 2,000 most frequent words on the same frequency list, or are of higher frequency than the items themselves.

Reflecting on the distribution of cognates, out of the 70 items on the ENVST, 31 were identified as non-cognate in either Romanian or Hungarian. This defined the number of cognates to be selected for the ROVST and HUVST respectively. Taking all three tests together, 93 out of 210 items are non-cognates, which accounts to 56% of items being cognates equally distributed on all three tests and levels. Furthermore, on the ENVST and ROVST 50% (19 and 20 cognates out of 39) are only English-Romanian cognates (bilateral). The other 50% are multilateral cognates. It has to be noted that this proportion does not represent the proportion of cognates in real language (addressed in Chapter 6), but the cognate distribution on the original VST. The HUVST’s cognate distribution slightly differs with 24 multilateral cognates versus 15 (bilateral) due to the lack of Hungarian words that are cognates in Romanian, but not in English (which also match the frequency level and selection criteria).

Moreover, the nouns (49), verbs (12), and adjectives (9) are equally distributed on the
three tests and, importantly, within each level. Although the test may seem heavily noun-bound, some English words can easily change POS (e.g. She looked upset [adj.]; the accusation upset him [verb]; this was the greatest upset [noun]) and as the VST is testing knowledge of word family, relying on Nation and Beglar (2007) it was assumed that learners are aware of the most common conjugations and derivations of the items. With regard to word length, it was also ensured that the number of syllables for each item is distributed proportionally on the three tests to arrive at a balanced set of items. Thus, the number of syllables varies from 1 (min.) to 6 (max.), with $M = 2.2$ on the ENVST, $M = 2.8$ on the ROVST and $M = 2.4$ on the HUVST. This shows that the mean number of syllables for each test was roughly equal.

Finally, all items on the new tests were put in a non-defining sentence. Both tests were checked by an additional educated NS in that language for cognateness, item-distractor relationship (i.e. that there are no distractors that could be misleading or equally acceptable as responses) and errors.

This chapter had two main aims. First, it presented the cultural and educational context, focalising on how vocabulary development is conceptualised in the Romanian context, and most importantly, pointed towards the paucity of a reliable frequency list and lexical tests that would allow teachers of Romanian to meaningfully explore the lexicon of learners of Romanian, or create graded readers in an informed way. Second, by addressing this gap, it synthesised the methodology for developing the RWL and its limitations, which led to the description and critical evaluation of the instruments utilised and developed for this thesis. The next section presents some of the criteria deployed to evaluate the results.
3.8 Issues of reliability and validity

Besides the definition of words and the consideration of different dimensions of vocabulary knowledge, another key issue is the meaningful and principled evaluation of the results elicited. This section thus focuses on the issue of testing vocabulary knowledge in a valid and reliable way. Reliability and validity are multi-faceted concepts that allow researchers to explore the ‘usefulness’ of tests (see Messick, 1989; Shepard, 1993; Weir, 2005) from an accuracy and consistency point of view on one hand, and whether the test is assessing what it purports to measure (Milton, 2009; Weir, 2005), on the other. It should be noted that, although, there are some exemplary studies (e.g. Fitzpatrick and Clenton, 2010; Fitzpatrick & Meara, 2004; Schmitt, Cobb, Horst and Schmitt, 2015; Schmitt, et al., 2001) that consider the employed and similar frequency-based lexical tests’ reliability and validity, more often than not, the tests’ practicality and simplicity plays a more important role in their popularity than their actual usefulness (Fitzpatrick and Clenton, 2010). Additionally, findings in the field can be characterised by the lack of replication studies that can either question, confirm, or refine corroborations in previous studies or verify their grounding in alternate settings (e.g. different language backgrounds, age groups), consistently leading to malignedit methodologies and generalisations (e.g. Schmitt, 2010).

The paucity of replications, sometimes even rigour, and the differing methodologies have perhaps led to the fact that it is still unclear, for example, how much vocabulary provides enough understanding or comprehension of spoken and written texts (lexical coverage) (Schmitt et al., 2015). Milton and Treffers-Daller (2013) assert that due to the different definitions of words and word selection procedures (dictionary sampling vs. counting units), different VS measures came to diverging results, meaning that the exact VS in English has been seriously overestimated for decades. Regarding word associations, Fitzpatrick et al. (2015) survey a number of word association studies from various disciplines and demonstrate that for a multitude of methodological reasons there are several
inconsistencies within this area of research and emphasise the lack and importance of norm lists that are evaluated with the target population in mind. The logical entailment of this is that, should the present studies be able to ameliorate the way multilinguals are seen from a vocabulary acquisition perspective, this could lead to more valid tests that can be reliably interpreted cross-culturally and perhaps even cognate norm-lists that can aid test development.

It is acknowledged that the process of validation is not definitive, whereby a test is deemed to be conclusively valid but rather as a continuous, evolutionary process (Chapelle, 1999; Weir, 2005). Consequently, prior to addressing the RQs, the instruments employed within the present study have to be explored through and evaluated against certain criteria. The evidence briefly discussed in relation to the findings is in light of the common understanding that tests ‘do not have reliabilities and validities, only test responses do’ (Messick, 1989: 14), therefore, these must be revisited for each instance of test use. Thus, despite the limited number of studies and participants involved, it is believed that elaborating on certain reliability and validity aspects of the results from a specific group of test-takers with a certain language background will enable exploring the findings more in-depth from a theoretical point of view, critically reviewing their potential, and highlighting their limitations.

Reliability is generally testing in statistical terms the degree to which the results obtained are stable over time, show a certain degree of consistency in terms of content sampling, and is free from bias (Weir, 2005). According to Weir (2005), inter alia, four different types of reliability testing, or in his terms, scoring validity can be distinguished. The first two types are the test-retest and the parallel forms reliability. Whereas the former one entails the same test being administered to the same participants within a reasonable time-frame, ideally before half a year after the initial test-time, the latter one requires a second test which is as close operationally to the original as possible. These two methods
establish reliability by comparing the scores between the two tests, and given the close scores obtained, one can conclude that the results are consistent and provide a reliable measure. The limitations of these tests lie in the method itself. The test scores in the second administration can be problematic as students’ test-taking and practice experiences can influence the scores, especially if an identical test is used, as memory may also have an effect (Weir, 2005).

The third type is called an equivalence measure and the fourth is marker reliability, which considers the consistency of scores with two or more markers. What is of interest here is the equivalence measure, which is employed to test for internal consistency of the test scores. The prerequisite criterion for this is that the items are homogeneous, meaning that they test the same ability. In this case, only one test administration is required and learners’ performance on one half of the items can be compared statistically to the other one (Weir, 2005), hence the name split-half reliability. The assumption here is that as items are selected to measure the same construct, test-takers’ scores (rank) should be approximately equivalent on both halves. Thus, a correlation of 1 indicates a perfect match and .7 suggests around 50% agreement. Within the current thesis, the conventional reliability coefficient of .8 is the acceptance level (e.g. Milton, 2009; Schmitt, 2010; Weir, 2005).

Another aspect of testing that ought to be addressed is construct validity. Vocabulary knowledge is a multi-faceted and complex construct, as the previous chapters have highlighted, and therefore, testing for all the different aspects of word knowledge is both challenging and impractical (Milton, 2009; Milton and Fitzpatrick, 2014; Schmitt, 2010). According to Schmitt (2010: 15), ‘the form-meaning link is the first and most essential lexical aspect which must be acquired, and may be adequate to allow recognition’. However, several studies suggest that the underlying construct that such tests measure might not allow a straightforward extrapolation on the items’ employability in production or meaning inferencing while reading, or on the depth of lexical mastery in terms of collocational or
derivational knowledge (e.g. Laufer and Goldstein, 2004; Kremmel and Schmitt, 2016). Thus, it is important to reiterate and stress that the VLT and VST are receptive vocabulary tests, based on a multiple-matching or -choice format, and the construct they attempt to measure is students’ ability to establish a form-meaning link. Weir (2005) argues that for validity, there are some a priori and a posteriori criteria that should be investigated.

Regarding the tests employed, the a priori criteria they follow, which can be also regarded as a theory-based validity, is that as learners acquire lexical knowledge incrementally, their accuracy should heighten at high-frequency levels and as frequency decreases, the scores should gradually diminish. Correspondingly, if scores confirm Meara’s (1992; see also Milton, 2009) theoretical profile of a language learner, this would also indicate satisfying content validity.

Content validity considers the items of a test and whether they all measure what they purport to measure. This is explored from a frequency point of view, but as in the case of multilinguals, lexical knowledge is not only frequency driven, the items will be also analysed from a cognate perspective. Furthermore, a valid lexical test should be able to distinguish between proficiency levels at different stages. In the present case, an objective external criterion is year of study at the university. Thus, first-year students should attain lower VS scores than their peers, and valid vocabulary tests should be able to capture this difference.

Finally, adding to the a posteriori validity check, as the participants are students, it is regarded that the external measure, or criterion validity, ought to be a variable that carries some meaning to teachers, test-takers, and educational stakeholders alike. It is a common practice to use an external measure that is either a standardised proficiency test (e.g. IELTS, TESOL) or representing general academic performance. As there is no standardised university entrance exam at the participants’ university, two academic performance indicators have been chosen: the end-of-year average when the test was taken and, in the case of students who completed three years, the average of these without the thesis mark
(Grade Point Average, henceforth GPA). The reason for this is that thesis marks are generally surmised to be inflated by various factors, such as supervisor input or the extended time to develop the thesis, and they also tend to focus on one specific area, in this case, literature. Thesis marks then seemed to be unfit for establishing concurrent validity. Nevertheless, both GPAs are weighted by credits of the modules, and consequently, they are regarded as a representative indicator of academic performance and a potential source for establishing predictive validity.
Chapter 4: Study One (S1)

Chapter 2 has highlighted that despite a wide body of research pointing to an undeniable facilitatory effect of cognates on not just learning, processing, and using a FL, but on vocabulary testing as well, research that systematically investigates a specific L2’s influence on English as an L3 in terms of VS is still missing. It also transpires that although there is a clear link between age, proficiency, and inter alia, (psycho)typological distance between learners’ background languages and the acquisition of a FL, no studies have compared VSs in two different languages, thus investigating the relationship between two. Furthermore, notwithstanding the facilitative effect of cognates even in languages that are not closely related on a lexical level, studies that employ a treatment-control paradigm to explore the impact of cognate instruction in different languages are a lacuna that needs to be embraced. In turn, this could substantially contribute to the current understanding of the lexicon and the ways in which learners’ attempt of growing their vocabulary might be expedited.

In light of this, this thesis aims to fill these gaps and investigate how English VS is influenced by a previously known Romance language in the case of Hungarian-Romanian bilinguals. The following sections thus present the RQs addressed, the methods involved, and the reasons for adopting and adapting them. Then, the focus turns to the design of S1, the description of participants and the cultural context. Following this, a detailed account is provided of the data analysis procedures, statistical packages employed, and methodological decisions taken. Finally, the results of S1 are presented, analysed, and interpreted.

4.1 Objectives and Research Questions

The aim of the present thesis is twofold. Firstly, in order to investigate the link between Romanian and English lexical knowledge overall and on the level of cognates, developing compatible tools that allow measuring vocabulary knowledge in Romanian as an L2 was
considered a prerequisite. Therefore, as Chapter 3 outlined, the research had to commence with creating a Romanian frequency list that serves as the basis of the new tests. As the vocabulary tests employed herein account for cognates in students’ DLC, I had to ensure that the newly developed tests match the design criteria of the original tests and once completed, the results are valid and reliable. Secondly, with these newly developed tests, I seek to examine how lexical proficiency in the L2 is comparable to L3 vocabulary size. That is, in the case of a Romance language and English, how are these languages integrated within the ML and in what ways they influence each other?

Additionally, since Nagy et al. (1993) indicate that raising awareness of cognates can have a facilitative effect, following Molnar’s methodology (2010), I became interested in whether the EI on cognates would result in an increase in the vocabulary scores of university students with high proficiency in English. Conversely, the recognition of cognates could be automatic (e.g. Hall, 2002), in which case, such instruction would not yield any significant differences between the groups.

In line with psycholinguistic traditions and in conjunction with the oft-repeated suggestion that obtaining detailed information about participants’ language background is essential (e.g. Treffers-Daller, 2011), I decided that each study will be complemented by a short questionnaire. This had two important functions. On the one hand, it enabled collecting detailed information on students’ languages and the age they started learning these. On the other hand, it allowed drawing a picture of students’ self-rated proficiency in the L2 and L3 that I could compare to the quantitative results (see e.g. Wei, 2007 for using a similar method to investigate the relationship between self-rated proficiency and vocabulary learning strategies). Finally, it provided insights into understanding how students perceive their L2 as a factor in learning English and their opinion on the cognate instruction.

In order to accomplish these objectives, the following RQs and hypotheses were formulated that are thought to serve as a systematic guide for the analysis and discussion
developed in this thesis:

- **RQ1**: What is the effect of explicit instruction on English-Romanian cognates on the vocabulary knowledge of English learners?

- **RQ2**: What is the relationship between overall lexical knowledge and cognate recognition in the L2 (Romanian) and L3 (English)?

The preceding chapters provided ample examples to support claims that strongly advise using cognates in teaching a FL (e.g. Morán-Molina, 2010; Rodríguez, 2001). There are several reasons for this. For one, cognates are regarded as obvious affordances that students can rely upon to infer meanings while reading or listening, which in turn facilitates fluency, and perhaps, even serves as a motivational factor for students (e.g. Hancin-Bhatt and Nagy, 1994; Nagy et al., 1993; Otwinowska, 2016).

Furthermore, raising awareness of the similarities between languages, especially on the level of lexis, can result in enhanced metalinguistic skills, increased test scores, and overall, a positive effect on LA in the case of learners at different ages and proficiency levels (e.g. Chen et al., 2012; Otwinowska, 2016). Nevertheless, the fact that not all cognates are recognised even after instruction or, conversely, the results that show that the instruction is beneficial, should be further investigated (e.g. Helms-Park and Perhan, 2016; Nagy et al., 1993; Otwinowska, 2016). RQ1 explores this.

In order to be able to compare the results to Molnar (2010), it was decided that the present study would also employ an experimental paradigm in which the Study (i.e. experimental) group (SG) receives EI on cognates. Even though Molnar’s study did not yield any statistically significant differences for the treatment group, due to the limited number of participants, different age and proficiency groups in the present case, it seemed logical for these results to be revisited. Therefore, it was hypothesised that due to the typological distance between Romance languages and English, Romanian L2 speakers could potentially capitalise on cognates and thus, an awareness-raising exercise focusing on English-Romanian structural similarities should result in higher scores for the SG.
RQ2 seeks to investigate the relationship between Romanian and English lexical knowledge from a number of perspectives. First, Molnar’s study (2010) found a correlation between students’ Romanian proficiency as rated by their teachers and VLT scores. Therefore, I regarded that the relationship between participants’ Romance L2 and English L3 can be further broken down and analysed at a finer level of granularity by employing compatible vocabulary tests. This way, lexical knowledge in the two languages is also explored from the perspective of cognates. It is hypothesised that since cognates are shared between the two languages, such words should provide an advantage or bridge for the participants that they can capitalise on, which in turn, the tests should be able to capture in a by-item analysis. Furthermore, it is also interesting to examine the interconnection between frequency, cognateness, and word recognition.

Second, due to the lexical similarities between the languages and the developed conceptual knowledge that proficient learners share, it is further hypothesised that lexically more proficient students in their L2 will also demonstrate more advanced knowledge in their L3. Under this, the hypothesis subscribes to and probes Cummins’ Interdependence Hypothesis described in section 2.7. In actuality, if there is a linear L2-L3 relationship on a lexical level, it is assumed that the association is even more accentuated for cognates, in comparison to non-cognates, and the tests should be able to reliably capture this.

Finally, if the above hypotheses are confirmed and languages are interconnected whereby cognates do have a considerable weight in language testing, it is entirely possible that certain items on the tests become less discriminative. Therefore, the reliability of the individual items will also be scrutinised, as it is considered that based on the similarities between participants’ background languages and English, future test development might be influenced. Put differently, prior lexical knowledge can influence the robustness of lexical tests and by inspecting the relationship between languages it becomes viable to tease apart a variety of factors affecting lexical assessments in the general sense that previously received
scarce attention.

The last part of the analysis takes this a step further and enables the triangulation of the empirical data elicited by the tests and the results of the questionnaire. As the questions focus on the age of onset of the L2 and L3, their self-rated confidence and proficiency levels in using the two non-native languages, and attitude towards cognates, I believe that this information should complement the results and allow for further investigating the lexicon from a student perspective. Since participants are considered Hungarian-Romanian bilinguals and grew up in Romania, it is probably safe to assume that students should have a larger vocabulary in Romanian, and due to the experience of using their L2 in various settings, they should also feel more confident and proficient in this language. Conversely, as Molnar’s (2010) study did not find any correlation between the age students started learning English and language proficiency, there is no clear expectation regarding this.

As an explorative question, the questionnaire also asked students to reflect on whether they find their Romanian knowledge helpful when learning English. Understanding students’ perceptions of the links between their FLs and the strategies they use was regarded essential, for various factors, such as students’ predisposition towards a language, psychotypology, and self-perceived proficiency, can all influence the development or use of a non-native language. Finally, I consider that in conjunction with the vocabulary tests, the questionnaire will provide invaluable data that can aid a more in-depth understanding of lexical knowledge in multiple languages within this cultural context.

4.2 Design of Study One

As previously indicated, the two studies presented in this thesis were developed to capitalise on the particular fact that cognates can provide learners with a substantial number of affordances in order to facilitate LL and that this is especially noticeable in the case of Romance languages and English. In order to investigate how Romanian as a Romance language, and an L2, influences English lexical knowledge, it was decided that measuring
Design of Study One

Lexical knowledge in both the L2 and L3 is crucial to gain a deeper insight into multilinguals’ ML.

Furthermore, to understand and evaluate cognate recognition, and explore whether this is an automatic process or proficient adult learners’ cognate knowledge can be enhanced by instruction on cognates, and the implications of this phenomena for VS tests, an experimental design was adopted. The experimental design allowed for measuring VS in both Romanian and English on different levels with participants being randomly allocated into a treatment (SG) and a Control group (CG). The reason for this was the necessity to establish a baseline level of lexical knowledge for the whole group, which was provided by the CG. The SG, as they belong to the same cohort, supposedly, possess (homogeneity of the population was ensured, see section 4.8.2) the same lexical and metalinguistic knowledge as the baseline group, which enables an inquiry into whether the EI is beneficial in terms of improving vocabulary scores overall or on cognate items.

Although it is necessary to acknowledge the fact that in the case of such experimental designs there is a potential problem of between-subject variability, it seemed reasonable that the experimental condition will reduce task-repetition effects that pose a risk if the same instruments were used after a longer treatment period (Milton and Meara, 1995). Moreover, even though longitudinal studies have several advantages, it was considered that two studies conducted in two separate years with different participants will provide a better picture of the relationship between Romanian proficiency and English lexical knowledge and allow exploring how the VLT and VST, and their Romanian versions respectively, can be utilised with multilinguals to account for cognate knowledge and investigate the relationship between a Romance L2 and English L3.

To summarise, following the development of the Romanian frequency list that enabled the development of Romanian frequency-based versions of the VLT and VST, participants in the case of both studies have been allocated randomly into either the CG or the SG. The
SG then received EI on the structural similarities between Romanian and English cognates. Subsequently, students filled out the pen and paper version of the VLT, followed by the respective Romanian test. The reason for choosing this order was the assumption that moving from the least known language to the more proficient one will keep students engaged in the study. Counterbalancing items on either frequency levels or between languages were surmised counterintuitive at the present stage.

As I am unaware of any VS studies that investigated lexical knowledge in different L2s and L3s, it was deemed that following the original test design, keeping items in high to low-frequency order, and languages separately, enables minimising unexpected effects that derive from the methodology employed. Additionally, the design already involves an experimental paradigm, hence any significant differences in scores would have been difficult to allocate to either the counterbalancing between languages and items, or the treatment condition. Finally, the entire questionnaire followed the vocabulary tests.

4.3 Ethical Considerations

Adhering to the Open University’s (OU) guidelines, prior to both studies, ethical approval was sought from the Human Research Ethics Committee and the research project had been granted a favourable opinion by Chair’s action as it was considered to be low risk. This means that the research protocol proposed for the current study has been accepted (HREC/2014/1626/Szabo/1).

Several ethical issues had to be considered though in the application. First of all, it had to be ensured that participants take part in the study voluntarily, were well informed of the purpose of the study, provided consent for their data to be used, and understood how this data will be used and disseminated in publications. Since all participants have already reached the age of consent, they were asked to sign a consent form (see Appendix 5) that allowed their data to be used anonymously. The consent of two lecturers, as gatekeepers, was also sought as the studies were conducted during formal teaching hours on university
Participants

In order to investigate the relationship between a Romance L2 and English L3, and address the proposed RQs, it was surmised that an L1 that offers the least possible lexical transfer to either Romanian or English was the most appropriate. Thus, the study was conducted at Partium Christian University in Oradea, Romania. This is a largely Hungarian university with most students and teachers being Hungarian NSs. The target group selected was enrolled in an English Language and Literature BA programme. This DLC also allowed a direct comparison of the results to Molnar’s (2010) study and focus on L3 acquisition within this cultural context.

Although more than 50 students were expected to participate in S1, due to availability at testing times, the actual number of participants who partook and whose results could be considered was reduced to 44. This number is distributed unevenly between first, second, and third-year students and due to unforeseen circumstances, it is also divided unevenly between the SG and CG. The reason for this uneven distribution in the experimental groups
Chapter 4: Study One (S1)

is that for random selection, it was decided that the study follows the way students attend seminars. For each seminar, students are randomly split in the beginning of their first year. This also allowed me to deal with the SG separately and answer questions on the EI, should they have any. Unfortunately, this resulted in that I could not actually control the number of participants who attended the seminar on the day.

Since the aim of the present study is to investigate students’ lexical knowledge, participants not scoring minimum 16 were not included in such analysis. The number of these students was four (two actually moved from Hungary to study at the university, therefore, they could not be regarded as Hungarian-Romanian bilinguals) and all of them belonged to the SG. Thus, the final number of participants is 40: 18 in the CG and 22 in the SG. Regarding year of study, age, and gender the numbers are distributed as follows: first-year \( n = 21 \), second-year \( n = 11 \), third-year \( n = 8 \); age ranges from 19 to 23 \((M = 20.48, SD = 1.32)\); with 31 females and 9 males. Participants reported that they started learning Romanian on average at the age of 5 \((min = 0, max = 8, SD = 2.4)\) and English at the age of 8 \((min = 2, max = 14, SD = 2.99)\).

Counterbalancing participants based on their L1 was also considered in line with psycholinguistic traditions. Ideally, testing Romanian NSs, who speak Hungarian as an L2 and are proficient in English as an L3 would have been revealing. As Molnar’s (2008) study showed, monolingual Romanians are likely to outperform both Hungarian monolinguals and Hungarian-Romanian bilinguals on an English vocabulary test (VLT), which suggests that the more proficient students are in a Romance language, the more likely they will be to demonstrate high proficiency in English lexical tests as well. Moreover, counterbalancing this way would have enabled, on the one hand, to see how Romanian monolinguals perform on the Romanian tests and compare this to Hungarian-Romanian bilinguals’ performance. On the other hand, it would have facilitated the analysis of any potential interaction between Hungarian L2 and English L3.
Unfortunately, finding participants with such a DLC, especially in formal settings, is quite challenging. Although Hungarians are required by the state to learn Romanian, and a couple of subjects in Romanian (Romanian History and Geography) throughout their education, this is not the case for Romanians. Thus, Romanian L1 speakers who also speak Hungarian either come from mixed-marriage families or from Hungarian families where they decided to educate the child entirely in Romanian. In this case, however, learners are very unlikely to receive any formal education in Hungarian. Obviously, there are exceptions to these categories, for instance, a Romanian family relocating to Hungary and learning Hungarian, or adults who either for work, family, or personal pleasure decide to learn Hungarian. In these cases, however, it is possible that chronologically and, maybe also in terms of proficiency, English is the L2, not Hungarian. Therefore, finding homogeneous participants who fit a Romanian L1, Hungarian L2, and English L3 criterion would have been time-consuming within the constraints of the present thesis. Furthermore, proficiency in the L2 and L3 should have been established prior to the study, which is problematic due to lack of such Romanian tools. Finally, perhaps due to age, educational or socio-economic background, within-subject variability would have yielded confounding results, and thus counterbalancing based on language proficiency was deemed impractical.

4.5 Instruments

Given the complexification in the lexicon by multilinguality and the interdependence of languages at the lexical level, it appeared reasonable to assume that exploring learners’ lexicon from a cognate perspective would not provide a realistic account if only students’ English vocabulary knowledge was tested. Simultaneously, despite the reliability of Yes/No tests, as cognates are not just about form, but their semantic information and the accurate recognition of these are equally important and under-researched, opting for a form-meaning mapping test format seemed to be justified. In light of these issues and the arguments outlined and critically evaluated in sections 3.6 and 3.7, the instruments employed here were
the Vocabulary Levels Test (VLT, Nation, 1983) and its Romanian version (RomVLT, for the tests’ layout and instructions, see Appendix).

### 4.6 Data Collection

The study consisted of three stages in total. Participants in the initial stage were divided into two groups for the experimental condition. As a first step, in order to be able to investigate the effect of Romanian-English cognate instruction and, similarly to Molnar (2010), the SG received a handout (Appendix 6) with EI on cognates and some examples. The handout was then discussed.

The second stage consisted of testing vocabulary knowledge in English and Romanian. The third stage, the post-test questionnaire (Appendix 3), as it were, followed straight after the two lexical tests. The three stages took approximately 40 minutes on average and I was present throughout in order to explain the aims of the study, sort out technical issues in the case they arose, answer questions, and ensure that students work individually without cheating. Finally, no participants reported any difficulties with understanding the tests, or visual or physical difficulties in completing the experiment.

### 4.7 Data Analysis

This section provides a review of the procedures followed at the data analysis stage. As the above avowed, the primary focus of the present thesis falls on the learner as a multilingual instead of merely on language- and testing-related aspects. However, to understand how multilinguals with a nearly identical shared language background perform on VS tests, and how their proficiency and cognate knowledge in one language influences VS in the other, I had to ensure first that the tools employed for such analyses are both valid and reliable. These issues are presented and reviewed in light of previous research within the ensuing section. Following this, the obtained results will be presented by addressing the formulated RQs systematically. Moreover, in order to facilitate comparisons and future
replications, it was deemed essential to make explicit references to Molnar’s (2010) and other relevant studies in the field.

The results obtained were entered into an Excel spreadsheet in a binary fashion, i.e. each correctly answered item received a score of 1, incorrect and non-attempted items were scored as 0. This dichotomous dataset allowed the separation of cognates from non-cognates, and also, the calculation of scores based on each frequency level. Thus, it was possible to use a cross-tabulation of cognates and non-cognates separated by frequency. Furthermore, this method enabled a by-item analysis in which the discriminative power or item facility index can be computed and compared to subjects’ ability estimates.

Since the development of the thesis was incremental in nature, it has to be said that the results of the first empirical study were initially analysed and graphically illustrated using SPSS. For the second study (S2), in order to gain a deeper understanding of statistics and allow for a more in-depth analysis, I opted for a statistical software that allows more freedom in terms of analyses and confirmatory visualisation. Therefore, the descriptive and inferential analyses for S2 (and at the write-up stage certain analyses for S1) were conducted in R (R Core Team, 2015). The pre-developed packages that were employed throughout the analyses are all approved by the Core Team (i.e. meet the scientific standard in terms of accuracy and descriptive material), are freely downloadable from the website, and include the following: \textit{ggplot2} for graphical representations, \textit{pastecs}, \textit{car}, \textit{reshape}, \textit{ez}, etc. for data-handling, descriptive, and inferential statistics.

In order to decide which type of statistical tests best fit the data, I ensured that the necessary assumptions for each parametric test are tenable. For reasons of consistency within the thesis, such analyses as the test for normal distribution, homogeneity of variance etc. are reported within the results section (4.8) together with their respective statistical analysis. Apart from correlations, \textit{t-tests}, and the traditional regression analysis including analysis of variance (ANOVA), in some instances, linear mixed-effects model analysis (also called
multilevel models or hierarchical linear models) have been adopted for the following reasons.

According to Jaeger (2008) mixed-effects models are well-suited for dichotomous datasets (i.e. binomially distributed variables: correct and incorrect responses), compared to ANOVAs have greater power, and allow the estimation of more accurate effect sizes. Furthermore, this type of statistical modelling allows including both within-subject and by-item variation in the same model, compared to having separate F scores. Moreover, it is not sensitive to violations of the assumptions of homoscedasticity and sphericity, and allows for testing the effects and interactions of both discrete and continuous predictors within one model. Finally, the fact that random intercepts and slopes can be included in the same model with subjects and items makes generalising the results of the experiment to other subjects and items in the population possible (see e.g. Field, Miles, and Field, 2012).

Accordingly, to address RQ1, paired sample t-tests were applied to the overall and cognate scores together with the group variable in the case of both languages to inspect the difference between the two groups.

Regarding RQ2, which investigates the L2-L3 relationship, the results on the Romanian and English vocabulary tests were compared. First, to explore the possible linear relationship (association) between the two scores, scatterplots were drawn. To further investigate this linearity, Pearson’s correlation coefficients were calculated between the two total scores and the two cognate scores. This enabled establishing the strength of the relationship between the two languages and a positive correlation value was an indicator of the fact that increased lexical proficiency in the L2 has the probability of resulting in increased proficiency in the L3. As the expectation is that the acquired L2, which is also the state language, is more proficient than the L3, which is a FL, paired-sample t-tests were conducted with their respective effect sizes to explore the magnitude of the relationship.

Following this, the results were investigated from a cognate facilitation perspective
and a succinct item analysis will be offered to explore whether background languages have
had an effect on the performance of certain items on the tests. The final aspect concerned the
relationship between proficiency and cognate scores. For this, mixed models were applied
to the data. This was a by-item analysis and for the sake of clarity, the variables used for
fixed and random effects, the procedure followed and, a posteriori assumptions are
explicated in conjunction with the results.

In order to determine whether the statistical tests employed in the current thesis are
significant or not, the traditional consensus employed within the field of Applied Linguistics
have been followed throughout (e.g. Field, Miles and Field, 2012; Weir, 2005). Thus, in the
case of correlation values between $\pm .3$ and $\pm .5$ were regarded a weak linear relationship,
between $\pm .5$ and $\pm .7$ were a moderate correlation, and above $\pm .7$ were interpreted as
a robust correlation. In the case of statistical significance and alpha values, the generally
accepted $p = .05$ were regarded as the cut-off point. Thus, any $p$-value below .05 was
considered as statistically significant and in line with educational research, two levels will
be reported $< .05$ and $< .01$. For effect sizes, $< .30$ were regarded as small, between .30 and
.50 were considered medium to large and $> .50$ were interpreted as large to very large effect
sizes (Cohen, 1988).

4.8 Results

This section is systematically organised around the RQs formulated in section 4.1 and
presents the analysis of the results pertaining to these questions. First, however, an analysis
regarding the reliability of scores and issues of validity are presented. Following this, the
experimental paradigm is explored through quantitative between-group comparisons to
reveal and analyse the effect of EI (RQ1). To address RQ2, a comparison is made between
the scores obtained on the VLT and RomVLT insofar as to investigate lexical knowledge
and cognate recognition. Finally, to further uncover students’ performance on the two
vocabulary tests and understand the relationship between their L2 and L3 and how they
perceive this, the findings will be compared to the questionnaire data using descriptive statistics and Spearman’s *rho* rank correlations. Students’ comments on the perceived benefits of their L2 and the EI on cognates will be qualitatively interpreted.

### 4.8.1 Issues of reliability and validity

Due to the issues raised in section 3.9 regarding the test-retest method (see Schmitt, 2010; Weir, 2005) or the feasibility of parallel testing methods, the two tests’ reliability and internal consistency were checked using the split-half analysis, which is regarded as the ‘industry standard’ (Weir, 2005: 30). The tests at first were split in two: 2k, 3k and 5k levels (high- and mid-frequency items) in one half and the 5k, 6k/UWL and 10k levels (mid- and low-frequency items) in the other one. In the case of the RomVLT, the alpha scores are .94 and .95 respectively, where the correlation between the forms equal .88. Cronbach’s alpha was calculated for individual items at $\alpha = .97$ and scores on the five different levels at $\alpha = .95$. For the VLT, the scores indicate that Cronbach’s $\alpha = .96$ for individual items and $\alpha = .95$ for the scores on different frequency levels. The split-half reliability test indicates $\alpha = .93$ in the case of the first half and .92 in the other, where the correlation between the forms equals .92. Since Cronbach’s alpha is above .90 for both tests, it appears that the reliability of the test scores are satisfactory.

Regarding the item statistics, it is important to explore two values – the inter-item correlations and item-total correlations. Inter-item correlation examines the relationship between scores to one item to scores on all other items in a scale, i.e. item redundancy or the extent to which items assess the same content (Cohen and Swerdlik, 2005). This value should ideally fall between the range of .20-.40, thus indicating that the items are, in actuality, reasonably homogeneous but also carry sufficient unique variance. In the case of the VLT this value is .22 and in the case of the RomVLT it is .30, both being within the expected range.

Item-total correlation is an indicator of the relationship between an individual item
score and the overall (total minus the actual item) score for the test. In simple terms, if a participant gets an item correct, it is expected that s/he will have a higher total score than others who provided an incorrect answer, hence the name item discrimination. This value is often regarded as the most fine-tuned measurement for finding out what participants know and which items can discriminate well between them. Generally, negative values would be a warning sign as this would indicate that a low score to that item results in a high overall score. Total-item correlations below .20 indicate low discriminatory power, between .20 and .39 indicate acceptable discrimination, and above .40 is regarded as very good discrimination (Field et al., 2012).

The total-item correlation values for the VLT range between .02 and .74, with an acceptable overall mean ($M = .48; SD = .16$) for all the items ($N = 84$ out of 90 as six items elicited a perfect score – discussion about this and additional item analysis is provided in the following section [4.8.3]). While eleven items are below the value of .30 and this suggests that such items should be reviewed, the average total-item correlation value for the items on the VLT seems satisfactory. With regard to the RomVLT, the correlation values are between -.09 and .84, with an acceptable overall mean ($M = .55; SD = .16$) for 87 items (three items had no variance due to perfect score). One item had a negative correlation value and further eight items exert low discriminative power ($< .30$), which is an indication of ambiguous or misfitting items. Overall, however, the results indicate that both tests seem to produce highly reliable scores with a significant level of internal consistency and satisfactory item statistics given the small to medium-sized population.

In order to establish whether the results obtained adhere to the validity criteria set in section 3.8, the data is explored in a number of ways. To begin with, descriptive statistics are used to present the overall scores and on the different frequency levels.

Prior to this, it has to be mentioned though that the total scores reported in the case of the VLT and RomVLT cannot be directly turned into VS scores as in the case of the VST.
This is because the VLT assesses knowledge on four selected frequency bands and the UWL, and thus to infer VS estimates out of the 10k most frequent words, the scores from these five bands would have to be turned into a figure that represents the number of words known, including the missing levels. On the one hand this is quite problematic (see Laufer and Ravenhorst-Kalovski, 2010) and on the other hand, as the results of tests are not used to compare students’ lexical knowledge in terms of reading coverage, but to explore their lexical knowledge in Romanian and English, it is considered that the raw total scores from the tests will provide relative-vocabulary size estimates that can be utilised to compare lexical knowledge in the two languages and across groups.

Table 4.1 reports the means and standard deviations on the different frequency bands and the total scores for the two tests. To reiterate, the difference between the two tests is that while the VLT features academic words from the UWL, on the RomVLT this was replaced by band 6 words.

### Table 4.1 Descriptive statistics for the VLT and RomVLT

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Total</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 5</th>
<th>Band 6/UWL</th>
<th>Band 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLT</td>
<td>40</td>
<td>59.43</td>
<td>15.18</td>
<td>15.20</td>
<td>10.93</td>
<td>11.93</td>
<td>6.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16.68</td>
<td>2.73</td>
<td>3.39</td>
<td>4.51</td>
<td>3.17</td>
<td>4.61</td>
</tr>
<tr>
<td>RomVLT</td>
<td>40</td>
<td>57.20</td>
<td>15.30</td>
<td>12.95</td>
<td>12.28</td>
<td>9.30</td>
<td>7.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21.49</td>
<td>2.99</td>
<td>4.56</td>
<td>4.19</td>
<td>5.63</td>
<td>6.07</td>
</tr>
</tbody>
</table>

First, the interest falls on the differences between the individual bands. It can be noticed that in the case of the VLT, the first two frequency bands (2 and 3) have almost equal means. In the rest of the bands, the scores suggest that students’ knowledge decrease from band 5 to 10 in comparison to the first two bands, except in the case of the UWL. By contrast, if the Romanian test is inspected, it can be noticed that as the frequency of items selected from the RWL decreases, the scores decrease as well. On band 3 and 5 the scores are very close to each other, however.
Figure 4.1 RomVLT mean results on different levels

Figure 4.1 summarises the mean scores on the different frequency levels on the Romanian test. As this also illustrates, students overall reached higher scores on the most frequent Romanian words than on the less frequent ones. This downward slope explains that the words chosen from the frequency list gradually become more difficult from left to right or from highly frequent to less frequent items.

Figure 4.2 VLT mean results on different levels
Figure 4.2 aids the comparison between the results of the two tests. In accordance with expectations, it demonstrates a strong overall frequency effect. Students attained relatively high scores on the 2k and 3k levels in the case of both languages and on lower frequency bands their scores diminish. This is in line with Milton (2007), who inferred that due to the strong effect of frequency on LL, tests that build on such frequency profiles retain a good construct validity. These results also resemble closely the theoretical profile of a learner’s lexical knowledge, as conceptualised by Meara (1992) and empirically demonstrated by a number of studies (see Milton, 2007, 2009; Schmitt et al., 2001; Schmitt, 2010). Therefore, overall the data suggests that there is some evidence towards construct validity.

To further investigate statistically the frequency effect, a multilevel regression model was fitted to the results using maximum likelihood ratio to establish the best fit. Constructing a multilevel model in the present case happened hierarchically. As a first step, using the scores that participants reached, which is our outcome or predicted variable, a baseline model (also called null model) is created, which contains the intercept only. Next, in order to account for the variation of scores across participants, a random intercept is added to the model. The difference between the two can be checked by inspecting the AIC (Akaike information criterion) and BIC (Bayesian information criterion) scores. However, as these values only serve as indication of a better fit, I will rely on the log-likelihood ratio (-2LL) which has a chi-square distribution and, therefore, it is reported in a similar fashion (Field et al., 2012). Following this, the predictor(s) are added to the model as a fixed effect. In this case, this enables establishing whether frequency is a significant predictor of the model, or in other words, whether the frequency of words had an effect on the scores.

Mixed-effects linear models also allow exploring whether the scores vary across not just individuals, but frequency levels as well. Therefore, the addition of a random slope to the model follows. If the difference between this model and the previous one, without the random slope, is significant, this shows that scores vary not just across individuals but within
frequency levels as well. Finally, interaction terms would be added to the model and if higher-order interactions are significant then the main effects are not interpreted. This brief explanation of the analysis served to understand the theory and procedure behind the multilevel models presented throughout the ensuing sections (see Field et al., 2012 for more details).

The aforementioned procedure resulted in four models that show a consecutively better fit. Thus, compared to the baseline model, adding a random intercept that accounts for variation across participants significantly improved the fit of the model, \( \chi^2(1) = 25.16, p < .001 \). This also suggests that using a mixed model instead of the traditional linear regression or ANOVA, which only accounts for fixed effects, is indicated. The third model includes the fixed effect, which is frequency levels in the present case. Similarly, the inclusion of a fixed effect resulted in a significant improvement \( \chi^2(4) = 250.49, p < .001 \). Finally, the addition of a random slope that accounts for the variation of frequency across participants also significantly improved the fit of the model, \( \chi^2(14) = 56.53, p < .001 \).

Calculating effect size for multiple regression models is not straightforward due to the complexity of the models. Some publications report no coefficient of determination at all, others report pseudo-\( R^2 \) or simply the adjusted-\( R^2 \) (\( R^2_{adj} \)) from the conventional linear model. However, both of these are problematic. For this reason, the technique I apply here is developed by Nakagawa and Schielzeth (2013), extended to random slopes by Johnson (2014), and is implemented in the piecewiseSEM package in R (Lefcheck, 2016). This technique produces two \( R^2 \) coefficients, one called Marginal (henceforth \( R^2_M \)), which includes only the variation explained by the fixed effects, and the Conditional (\( R^2_C \)), which takes into account both fixed and random effects (for a more detailed discussion on the advantages of this technique, see Johnson, 2014). For the present model, thus, \( R^2_M = .45 \) and \( R^2_C = .97 \), which means that the fixed effects explain 45% of variance in the model, which is always exactly the same as if a traditional regression model was fitted to the data, and by
accounting for the random effects, the variance explained by the model significantly increases. The final model is provided in the following table.

### Table 4.2 Model of VLT scores by levels

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>$\beta$</th>
<th>SE</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>15.18</td>
<td>0.43</td>
<td>35.21</td>
<td>***</td>
</tr>
<tr>
<td>Eng 3k</td>
<td>0.03</td>
<td>0.30</td>
<td>0.08</td>
<td>.93</td>
</tr>
<tr>
<td>Eng 5k</td>
<td>-4.25</td>
<td>0.44</td>
<td>-9.70</td>
<td>***</td>
</tr>
<tr>
<td>Eng UWL</td>
<td>-3.25</td>
<td>0.31</td>
<td>-10.36</td>
<td>***</td>
</tr>
<tr>
<td>Eng 10k</td>
<td>-8.96</td>
<td>0.53</td>
<td>-16.84</td>
<td>***</td>
</tr>
<tr>
<td>Random Effects</td>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng 3k</td>
<td>1.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng 5k</td>
<td>2.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng UWL</td>
<td>1.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng 10k</td>
<td>3.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sig. codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

As Table 4.2 indicates there is a significant relationship between the obtained scores and the frequency levels. First, it can be noticed that words from the 3k level did not significantly predict 2k level words, $b = 0.03, t(156) = 0.08, p = .93$. In other words, there is no significant difference between the two levels. Regarding the lower frequency bands and the UWL, each level is a significant predictor of the model and the negative $b$ values indicate that vocabulary scores on the different levels decrease overall and significantly. The standard deviation scores explain that across levels and individuals the scores vary most on the 10k band. It has to be noted that as for this model only one predictor or independent variable was used, the model cannot be affected by multicollinearity. Furthermore, this model could be further built to see how the age of acquisition or proficiency, for example, improves the model. This, however, will be investigated in the following. For now, the Romanian scores will be investigated regarding the frequency of items and whether this is a similarly good predictor of vocabulary scores as on the English test.

The Romanian model was built following exactly the same procedure as described
above. The results stipulate that the scores are best explained by allowing these to vary across frequency levels and participants: $X^2(4) = 159.65, p < .001$. In the next step, I added frequency as a fixed effect, which emerged as a significant predictor of the model. The final model $X^2(14) = 96.41, p < .001$ explains a variance of 26% based only on the fixed effects as $R^2_M = .26$, and $R^2_c = .97$ including the random effects.

Table 4.3 indicates that compared to high-frequency words, the scores decrease significantly on each frequency level as indicated by the negative $b$ values. Compared to the English test, it also demonstrates that the items from the 3k level also become a significant predictor of the model.

### Table 4.3 Model of RomVLT scores by levels

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>$\beta$</th>
<th>$SE$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>15.30</td>
<td>0.47</td>
<td>32.38</td>
<td>***</td>
</tr>
<tr>
<td>Rom 3k</td>
<td>-2.35</td>
<td>0.42</td>
<td>-5.61</td>
<td>***</td>
</tr>
<tr>
<td>Rom 5k</td>
<td>-3.02</td>
<td>0.37</td>
<td>-8.15</td>
<td>***</td>
</tr>
<tr>
<td>Rom 6k</td>
<td>-5.98</td>
<td>0.66</td>
<td>-9.07</td>
<td>***</td>
</tr>
<tr>
<td>Rom 10k</td>
<td>-7.95</td>
<td>0.77</td>
<td>-10.30</td>
<td>***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.77</td>
</tr>
<tr>
<td>Rom 3k</td>
<td>2.18</td>
</tr>
<tr>
<td>Rom 5k</td>
<td>1.81</td>
</tr>
<tr>
<td>Rom 6k</td>
<td>3.85</td>
</tr>
<tr>
<td>Rom 10k</td>
<td>4.60</td>
</tr>
<tr>
<td>Residual</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Sig. codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Turning back to the validity issues that are of concern within this section, theoretically, it can be argued that the tests measure an underlying trait that is in a close approximation of what happens in real life: as the items’ frequency decreases, vocabulary knowledge diminishes as well. However, in the case of content validity, it can be noticed that the mean results of the RomVLT follow the expected frequency effect, albeit the English test indicates some discrepancies.

This discrepancy is evident in the case of the UWL, as the scores do not align with the
decreasing tendency (Figure 4.2). The reason for this is twofold. On the one hand, it can be argued that as the participants are undergraduates, they have a relatively high exposure to academic words and thus, it can be speculated that they acquire them more easily. The other reason is that almost all words (17 out of 18) on this level are cognates, and consequently, participants could rely on their Romanian knowledge to work out, or indeed transfer, the meaning of these items.

This potential transfer can be regarded as evidence for the importance of determining and taking into account students’ background languages when testing the lexicon of multilinguals. By way of instantiation, if the results obtained in the current study were to be directly compared to Molnar’s (2008) monolingual Hungarian participants, the content validity of the two tests would make it problematic to meaningfully interpret the scores and compare the results, as those participants were younger in age, and were neither university students, nor Romanian speakers to be able to rely on cognates. Thus, the test would not perform uniformly across cultures and proficiency levels (see Chapter 6 for the implications of this).

Another issue that has been deemphasised so far is that the first two bands of VLT do not discriminate between students’ overall VS. Although it is somewhat expected that English undergraduates should be able to recognise a reasonably high percentage of the 3k most frequent words, the items should still be able to discriminate between students of different abilities at these levels. The performance of the items, in terms of item facility and discriminative power, is further analysed in the subsequent sections. Consequently, while there is some evidence for construct validity, in the case of Romanian L2 speakers and university students, the VLT’s content validity cannot be taken at face value.

It is also expected that as students progress in their studies, their vocabulary knowledge also improves. As a valid vocabulary test, the VLT should be able to capture this lexical growth and reveal such differences.
Table 4.4 Descriptive statistics of the two tests by year of study

<table>
<thead>
<tr>
<th>Test</th>
<th>Year of Study</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLT</td>
<td>1</td>
<td>21</td>
<td>51.86</td>
<td>16.57</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11</td>
<td>66.00</td>
<td>12.02</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>70.25</td>
<td>13.63</td>
</tr>
<tr>
<td>RomVLT</td>
<td>1</td>
<td>21</td>
<td>49.24</td>
<td>21.56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11</td>
<td>66.73</td>
<td>18.11</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>65.19</td>
<td>19.33</td>
</tr>
</tbody>
</table>

As Table 4.4 above confirms, based on the mean scores, vocabulary knowledge increases across the three academic years. As the random effects in the model did not yield a significant improvement, using mixed models results is exactly the same coefficients as the conventional linear models. Therefore to check the difference between years of study, separate ANOVAs were conducted for the English and Romanian test results and this was then broken down applying independent \textit{t}-tests with Bonferroni-adjustments. The Shapiro-Wilk test confirms ($W = 0.97, p = 0.39$) that the assumption of normality regarding the VLT is tenable, and the performed Levene’s test confirms that the assumption of homogeneity of variance between the years is also tenable ($F(2, 37) = .85, p = .43$). The ANOVA explains that the differences in the means are significant ($F(2, 37) = 5.88, p = .006; \text{R}^2_{\text{adj}} = .20$) and the model explains 20% of the variance. There is an increase from year one to year two of 14.14, which is significant at the .05 level (using Bonferroni-adjustment $p = .045$) and an increase of 18.39 to year three (Bonferroni-adjusted $p = .016$). A linear trend analysis also confirms that there is a significant linear trend $t(37) = 2.97, p = .005$ with a medium effect size ($R^2 = .44$), indicating that as the year of study increases, VLT scores increase as well.

Regarding the Romanian test, although there was no expectation in terms of how well the test differentiates between years of study as not all students study Romanian formally in the same way as English, the mean scores indicate that there is a difference between year one ($M = 49.24, SD = 21.56$) and the other two years ($M = 66.7$ and 66.73; $SD = 18.11$ and 19.33 respectively). Since normality can be assumed, based on the Shapiro-Wilk test ($W = .95, p = .07$) and Levene’s test confirmed homogeneity of variance ($F(2, 37) = .49, p = .64$)
as being tenable, an ANOVA was conducted to check whether there is a significant effect. This confirmed that there is a small, but significant effect $F(2, 37) = 3.43, p = .042$, which explains 4% of the variance ($R^2_{adj} = .04$), however, the Bonferroni-adjusted $t$-tests revealed non-significant differences between the years.

It can be inferred from these results that the VLT performs according to the expectations and it can detect differences in vocabulary scores which are in line with an expected proficiency increase as students academically progress. However, as participants are not formally studying Romanian, the scores are not expected to increase with years of study and this is also reflected.

In order to explore whether there is a linear relationship between the size of testees’ lexicon defined by the overall test scores and the two academic performance indicators, first, Spearman’s rank correlations were utilised since these violated the assumption of normality. The results indicate that VLT results correlate moderately with the average marks in the year when the test was taken: $\rho = .65, p < .001, N = 38$ (the number of data points is different here as not all subjects reached the end of the academic year or graduated from university). Similarly, there is a moderate correlation between the English scores and GPAs at the end of year three: $\rho = .60, p < .001, N = 31$. The Romanian test, however, indicates a weaker correlation between the test scores in the case of the end-of-year results ($\rho = .47, p = .003, N = 38$) and the GPA ($\rho = .39, p = .03, N = 31$) as well. These relationships explain that VS measures in both FLs can be a good indicator of academic performance. However, due to the fact that all students study English, the VLT constitutes a seemingly better predictor.

Finally, to address the predictive validity of the results as described in section 3.8 and investigate the relationship between academic achievement and VS, a regression model was fitted to the data, where GPA (as the correlation between the two indicators was robust $r = .93$, only the end of year values were taken as these contain less missing values, $N = 38$) was used as outcome value and the English and Romanian test results as predictors. Three
Results

separate models were built. In block one, GPA was predicted only by the VLT scores: \( F(1, 36) = 28.5, p < .001 \), \( R^2_{adj} = .43 \). This can be interpreted as the VLT is a significant predictor of academic performance as indicated by the GPA, and the test scores explain 43% of the variance. In the second block, only the RomVLT results were used as predictors, and this yielded a significant model as well: \( F(1, 36) = 10.05, p < .001 \), \( R^2_{adj} = .20 \). The third block included both tests and their interaction in the model. Despite the fact that the model overall is significant, \( F(3, 34) = 9.68, p < .001 \), \( R^2_{adj} = .41 \), neither the interaction \( t(34) = -0.7, p > .05 \), nor the RomVLT \( t(34) = 0.30, p > .05 \) had a significant effect on the model. Therefore, it appears that the VLT is a significant predictor of academic success and the larger students’ vocabulary is, the higher their GPA is likely to be, taking into account other latent factors that the model cannot account for.

In conclusion, the overall results indicate some evidence for the validity and convincing evidence for the reliability of the data in the case of both tests. Some further investigations regarding the validity of the test will be afforded in the forthcoming sections. However, in light of the above evaluation and findings, I now turn to the experimental paradigm and the comparison of lexical knowledge in Romanian and English, bearing in mind the proposed RQs.

4.8.2 The effect of explicit instruction (EI) on cognates (RQ1)

The first step in the main analysis of the results is to explore the effect of the experimental treatment. The primary concern herein is to investigate whether the awareness-raising exercise influenced the SGs’ scores, compared to the CG that received no EI on the structural similarities between the two languages.

Prior to this however, to lend reasonable confidence to the fact that the two groups are homogeneous at the start and the effect of experimental treatment is not confounded by some extraneous factors, the equality (homogeneity) of the groups should be established. In the lack of pre-tests, this can be rectified by other objective indicators, which in this case were
the GPA and age of acquisition. Since the GPA and age variables violate the assumption of normality, the non-parametric Wilcoxon rank-sum test with continuity correction was utilised (this test is equivalent to the Mann-Whitney U test). According to this, equality between the two groups can be assumed based on the GPA ($W = 277, p = .16$; SG: $M = 8.41, SD = 1.39, n = 21$; CG $M = 7.93, SD = 1.14, n = 17$), the age of onset of learning English ($W = 170, p = .45$; SG: $M = 7.45, SD = 2.36, n = 22$; CG $M = 8.50, SD = 3.60, n = 18$), and Romanian ($W = 234, p = .32$; SG: $M = 5.64, SD = 2.11, n = 22$; CG $M = 4.67, SD = 2.70, n = 18$). This means that the two groups potentially include learners from the same population, which also entails that should the EI on cognates have an effect on the SG, this can be potentially ascribed to the treatment condition and should show accordingly in their scores.

![Figure 4.3 Barplot for the Study and Control groups on the two tests](image)

Figure 4.3 indicates that the group that received the instruction on cognates performed on average slightly better than the CG on the VLT. However, on the RomVLT it appears that the reverse is true, the SG actually attained a lower overall score than the CG.
Table 4.5 Descriptive statistics for the Study and Control groups on the two tests

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLT</td>
<td>Study</td>
<td>22</td>
<td>60.59</td>
<td>17.26</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>18</td>
<td>58.00</td>
<td>16.32</td>
<td>29</td>
</tr>
<tr>
<td>RomVLT</td>
<td>Study</td>
<td>22</td>
<td>52.86</td>
<td>23.08</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>18</td>
<td>62.50</td>
<td>18.63</td>
<td>34</td>
</tr>
</tbody>
</table>

The scores reported in Table 4.5 show some indication that the SG performed slightly better than the CG, with similar consistency on the VLT. For the assumption of normal distribution is met as the Shapiro-Wilk test indicates (Study: $W = 0.96, p = .50$; Control: $W = 0.97, p = .83$) and the assumption for homogeneity of variance between the groups is also tenable as demonstrated by Levene’s test ($F(1, 38) = 0.038, p = .91$), hence, a two-sample t-test was employed to investigate the difference. This confirms that the means are equal (non-significant) with a small effect size ($R^2 = .08$).

In the case of the Romanian test, this effect, or rather the lack of it, seems to be reversed as the CG ($M = 62.50, SD = 18.63$) outperformed the SG ($M = 52.86, SD = 23.83$). Although the necessary assumptions for normality (Study: $W = 0.93, p = .14$; Control: $W = 0.93, p = .21$) and homogeneity of variance ($F(1, 38) = 1.16, p = .29$) have been met, the t-test indicates that the difference between the means is not significant, with a small effect size ($R^2 = .23$).

Table 4.6 Cognate results on different frequency levels on the VLT

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Cog Total</th>
<th>Cog 2k</th>
<th>Cog 3k</th>
<th>Cog 5k</th>
<th>UWL</th>
<th>Cog 10k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>22</td>
<td>35.64</td>
<td>8.82</td>
<td>9.00</td>
<td>3.50</td>
<td>11.45</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.35</td>
<td>1.44</td>
<td>1.88</td>
<td>1.41</td>
<td>2.89</td>
<td>1.83</td>
</tr>
<tr>
<td>Control</td>
<td>18</td>
<td>35.83</td>
<td>8.78</td>
<td>9.06</td>
<td>3.44</td>
<td>11.61</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.06</td>
<td>1.06</td>
<td>1.16</td>
<td>1.20</td>
<td>3.03</td>
<td>1.80</td>
</tr>
</tbody>
</table>

Although the mean scores confirm that the instruction on cognates had no effect on the scores, to assure consistency, the cognate scores will also be investigated severally on the different levels. As the instruction was on cognate relationships, these are the most probable to indicate an effect, if one exists.

Table 4.6 reiterates the fact that the awareness-raising exercise on cognates did not
enhance the SG’s performance as the means are virtually identical overall and on the
different frequency bands as well. Despite the slightly higher means for Romanian cognate
recognition in the case of the CG as confirmed by Table 4.7, a *t-test* indicates that there is a
non-significant difference between the two means.

Table 4.7 Cognate results on different frequency levels on the RomVLT

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Cog Total</th>
<th>Rom2k</th>
<th>Rom3k</th>
<th>Rom5k</th>
<th>Rom6k</th>
<th>Rom10k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>22</td>
<td>28.23</td>
<td>8.36</td>
<td>5.86</td>
<td>3.45</td>
<td>8.32</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.98</td>
<td>1.56</td>
<td>2.95</td>
<td>1.22</td>
<td>5.45</td>
<td>1.95</td>
</tr>
<tr>
<td>Control</td>
<td>18</td>
<td>32.44</td>
<td>8.72</td>
<td>7.28</td>
<td>4.00</td>
<td>9.78</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.69</td>
<td>1.23</td>
<td>2.22</td>
<td>0.84</td>
<td>5.16</td>
<td>2.33</td>
</tr>
</tbody>
</table>

To conclude RQ1, the results suggest the EI on cognates had no effect either on the
total or cognate scores. One interpretation is to assume that this is due to the fact that students
were randomly assigned to groups and therefore, it is entirely possible that the CG’s lexical
knowledge in English is slightly better, albeit this is not supported by the GPAs. The other
position is that perhaps, in conjunction with the conclusions obtained by García (1991) and
Nagy et al. (1993), due to students’ age and language proficiency, cognate recognition has
become an automatic process and as a result, the effect of the instruction goes unnoticed.
These interpretations are further investigated in the last sub-section and discussed in Chapter
6. However, thus far it can be concluded that the tests did not prove that the EI on cognates
helped students with their lexical performance in English. Furthermore, it can also be
inferred that the two groups are identical in terms of lexical knowledge, age of acquisition,
and GPA, and consequently they belong to the same population. Therefore, for the following
analyses, the two groups will be treated as one (N = 40).

4.8.3 Exploring the relationship between the English and Romanian lexica (RQ2)

To systematically address RQ2, the results that pertain to this sub-section will be
structured into three comprehensive parts. Part A explores the overall results in a within-
subject analysis. In particular, it compares vocabulary knowledge in participants’ L2 and L3
Results

based on the total scores and, then, on cognate scores, followed by the investigation of the relationship between overall vocabulary knowledge and cognate knowledge from a multilingual perspective. Part B employs a by-item analysis to specifically address the cognate facilitation effect. Finally, part C explores the effect of lexical proficiency on the scores taking into account cognateness and word frequency.

A. Comparing vocabulary size in English and Romanian

In order to compare lexical knowledge in participants’ L2 and L3 based on the elicited scores, I commence by exploring the existence of a linear relationship between the overall scores. As Table 4.1 shows, the mean scores on the two tests overall are very close to each other with slightly higher scores on the English test. In addition, according to the standard deviation scores, participants on the English test were slightly more consistent ($SD = 16.86$ versus 21.49) with a minimum score of 22, compared to 16 on the RomVLT.

The graph below (4.4) illustrates the robust correlation that obtains between the overall scores. The closer students’ results are on the two tests, the closer they are to the equation line. To put it differently, the stronger the association is between the two tests, the better the relationship between English and Romanian lexical knowledge can be described. This is supported by the significant correlation between the two tests as Pearson’s $r = .79$ ($p < .001$, two-tailed in all cases). The individual frequency bands show strong correlations as well: for the 2k levels $r = .64$; 3k levels $r = .63$; 5k levels $r = .64$; 6k and UWL level $r = .82$, and the 10k levels $r = .72$.

If the scores are separated into cognate and non-cognate items, the correlation coefficients become $r = .87$ and .65 respectively. This indicates that cognate scores show a more robust linear relationship in the two languages than non-cognate scores, and cognate knowledge in one language is a strong predictor of cognate knowledge in the other. Moreover, the correlations show that cognate knowledge in the case of English is a better indicator of overall Romanian knowledge ($r = .84$) than non-cognate knowledge ($r = .72$).
Similarly, Romanian cognates are a better predictor of English VS ($r = .83$) than non-cognate knowledge ($r = .72$). Consequently, it can be said that cognate knowledge in one language is a very strong predictor of not just cognate knowledge in the other, but overall lexical knowledge in both the L2 and L3.

Following this, regression analyses were carried out to gain more insights into the L2-L3 relationship. First, the interest fell on whether overall lexical knowledge in Romanian is a significant predictor of lexical knowledge in English. This led to a significant model: $F(1, 38) = 62.71, p < .001, R^2_{adj} = .61$, which explains that the RomVLT scores account for 61% of the variance. Next, the predicted variable was kept as the total VLT scores, however, the Romanian scores were separated into cognate and non-cognate scores, to explore whether either of them or the interaction between the two results in a better model. Although this model was significant, the interaction between L2 cognate and non-cognate scores was not
significant, therefore, it was removed from the model. The final model thus became: $F(2, 37) = 47.4, p < .001, R^2_{adj} = .70$. The fixed effects, however, indicate that only the L2 cognate scores are significant predictors of the model $b = 1.98, t(37) = 5.09, p < .001$, non-cognates are just above the accepted significance level $b = -0.80, t(37) = -1.99, p = .054$.

Additionally, due to the high correlation between L2 cognate and non-cognate scores, this model violates the assumption of multicollinearity as the Variance Inflation Factor (henceforth VIF) for the predictors is above 5 and the tolerance value is below .2. Therefore, due to multicollinearity and the non-significant effect of non-cognates on the model if used in conjunction with cognates, this variable was removed from the model, which led to $F(2, 38) = 84.27, p < .001, R^2_{adj} = .68$, where the predictor is $b = 1.25, t(38) = 9.18, p < .001$. Subsequently, if this model, which uses only L2 cognate scores and explains 68% of variance, is compared to the overall L2 scores, which explain only 61%, it seems Romanian overall and cognate knowledge both predict L3 scores, albeit cognate scores alone explain more variation, and therefore, result in a better fit.

Adhering to exactly the same procedure, but this time using the Romanian scores as the predicted variable, also revealed similar effects. The results indicate that both the English cognate scores ($F(1, 38) = 89.58, p < .001, R^2_{adj} = .69$) and non-cognate scores ($F(1, 38) = 89.58, p < .001, R^2_{adj} = .50$) are a significant predictor of L2 knowledge, but again, cognates explain more variance in the model. In conclusion, it can be inferred that there is a significant relationship between the L2 and L3, and this is even more accentuated in the case of cognates. This positive relationship also suggests that the larger the vocabulary knowledge in one language, the larger it is in the case of the other language, especially with regards to cognates. If this is generalised, it appears that in the case of typologically proximate languages beyond the L1, such as Romance languages and English, cognates can serve as a bridge between the languages. This will be further discussed in Chapter 6.

One additional issue should be addressed regarding the overall scores. The slightly
higher results on the English test are somewhat unexpected, if the fact that Romanian should be participants’ L2, both chronologically and in terms of proficiency, since they live in Romania, is taken into account. Before an inquiry into this difference, it has to be ensured that the necessary assumption for *t*-tests is met by the data. In order to check for normal probability distribution in the case of the total scores on the VLT and RomVLT, the Shapiro-Wilk normality test was carried out. The test indicates that in the case of both tests the assumption for a normal distribution is tenable, $W = 0.97$, $p = .40$ for the English test and $W = 0.95$, $p = .73$ for Romanian. The paired-samples *t*-test indicates that the difference between the two scores is non-significant: $t(1, 39) = 1.07$, $p = .29$. In other words, contrary to expectations, participants’ performance in the L2 and L3 is identical as the equal means suggest. Whether this is a repercussion of the two tests employed herein (academic items on the VLT or the effect of cognates) or a true representation of students’ proficiency, will be further explored in S2 and discussed in Chapter 6.

**B. The effect of cognate facilitation on testing**

The second issue that pertains to the relationship between the L2 and L3 is the effect of cognates on lexical tests due to similarities between the languages. In order to determine whether cognates are more easily recognised than non-cognates, the percentage of correct answers to cognate and non-cognate items is considered in the first instance.

**Table 4.8 Descriptive statistics for cognate and non-cognate items**

<table>
<thead>
<tr>
<th>Test</th>
<th>Cognate Status</th>
<th>N</th>
<th>Mean Item Facility in %</th>
<th>Item facility range</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>Min</em></td>
<td><em>Max</em></td>
</tr>
<tr>
<td>VLT</td>
<td>Cognate</td>
<td>48</td>
<td>74.43</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>NC</td>
<td>42</td>
<td>56.43</td>
<td>12.5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>90</strong></td>
<td><strong>66.03</strong></td>
<td><strong>10</strong></td>
<td><strong>100</strong></td>
</tr>
<tr>
<td></td>
<td>Cognate</td>
<td>48</td>
<td>62.76</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>NC</td>
<td>42</td>
<td>64.46</td>
<td>17.5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>90</strong></td>
<td><strong>63.56</strong></td>
<td><strong>17.5</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 4.8 above summarises the proportion of correct answers to cognates and non-
cognates. The reported figure is called item facility index or the difficulty of an item. The item facility index of cognate and non-cognate scores, demonstrates that on the VLT to 48 cognate items the percentage of correct answers is over 74%, whereas to non-cognate items it is 56%. In contrast, on the RomVLT the percentages are not only closer to each other but slightly more non-cognate items were answered correctly.

To analyse whether this difference is substantial, first, the normality and homogeneity of variance of the variables need to be checked. The Shapiro-Wilk normality test is significant in the case of both variables, meaning that the assumption of a normal distribution is violated. Although in this instance an appeal could be made to the Central Limit Theorem as group sizes (cognate/non-cognate items) are larger than 30 and thus parametric tests may be equally used, since the two tests largely produce congruent results, the standard procedure is adopted (Field et al., 2012). Therefore, to compare the mean percentage of correct answers to cognates and non-cognates, the non-parametrical Wilcoxon rank-sum test was applied. In the case of the VLT, the difference is statistically significant ($W = 1410, p < .001$). However, it is non-significant for the RomVLT ($W = 954, p = .67$).

Since cognates in English are of a lower frequency and, therefore, more common words in Romanian, the results indicate that this provided a possible source of cross-linguistic transfer that participants could capitalise on in English, but not in Romanian. If this is further explored on the individual frequency bands, there is a noticeable difference in the percentage of correct answers on band 4 ($M = 66.69$ for the UWL and $M = 51.81$ for the 6k words on the Romanian test). While for the VLT these words are sampled from the UWL, in Romanian they derive from the 6k most frequent words, thus this difference means that the UWL provides 17 cognates for test-takers that are also more familiar to university students, while the Romanian test, although, presumably the 17 cognates should be roughly from the same frequency, these are not restricted to academic words. This again, can be regarded as evidence for the importance of accounting for test-takers’ language background,
the number of potential cognates, and also their types, when it comes to lexical tests.

Another issue that is worth investigating is the fact that the VLT’s first two bands do not facilitate discriminative power, as mentioned previously in the validity section. Additionally, Table 4.8 indicates that the maximum percentage of correct responses was 100, which explains that one or more items elicited a ceiling effect. In order to explore whether this issue is due to cognate knowledge or some other discrepancy, a brief item analysis is presented. First, however, it should be mentioned that no ceiling effects were present with any of the testees in English, ability estimates (total score divided by number of items) ranged from .24 to .99 (\( M = .66; \ SD = .19 \)). On the Romanian test, one student reached a perfect score (1 = 90 items) and the least able student’s ability estimate was .18 (\( M = .64; \ SD = .24 \)). Item difficulties on the VLT ranged from .10 to 1 (\( M = .66, \ SD = .27 \)) and on the RomVLT from .18 to 1 (\( M = .64, \ SD = .22 \)).

Initially, the English items that reached a ceiling effect are inspected. There were six such items in total: total, victory (2k); angel, darling, illustrate (3k); and anonymous (UWL). This reveals that except for darling, all of the items that elicited a perfect score are cognates. Although there are no items that elicited a floor effect, it is worth analysing items with an item facility index of less than .25, as these are considered to be too difficult for students, which means that they do not discriminate well amongst their abilities. In total there were eight of these items: indigenous and frustrate (UWL); dabble, acquiesce, crease, skid, dregs, and casualty (10k). Interestingly, the two items from the UWL are both cognates. The potential reason for students’ miss on these items is due to ambiguous distractors as the analysis reveals. In the case of indigenous, the definition was native, however, one distractor was maternal which is more likely to be known by students due a semantic link, even if this is not too obvious in English: mother tongue in Romanian is translated as limbă maternă. This carries a strong semantic link between native and maternal for speakers of both languages.
In the case of *frustrate* the anomaly is not as obvious. The incorrect answers that students chose were *coincide* and *expel* for the definition ‘*to prevent someone from doing something they want to do*’. *Coincide* is a cognate, which means that the meaning in the two languages overlaps, thus there is no substantial evidence to interpret why students were misled by this distractor. With regard to *expel*, one can assume the logical cause-effect relationship that if someone is expelled, they will be prevented from going back, which provides some evidence why the two items could be ambiguous, albeit in this case, the reason is not cognateness.

The six items with very low hit rate from the 10k level are all non-cognates and as a general observation, it can be inferred that these are non-response items rather than incorrect ones. While marking the tests, it emerged that many students did not even attempt guessing these lexemes as in many cases 10k items were left simply blank. The reason for this is that on the one hand, the majority of items on this level are non-cognates. On the other hand, these are low-frequency items that are expected to be known only by students with a larger vocabulary, which is a good sign of validity.

In the case of the RomVLT, as the aforementioned has pointed out, the first two levels differ in the mean scores, suggesting that as items become infrequent, they are more difficult for students, and therefore, result in a lower hit rate. Three items resulted in a ceiling effect all of them cognates from the 2k band: *asistent, repeta, and stabilit* (*assistant, repeat, and established*). Items that proved to be the most difficult (facility index below .30) are all from the 10k band: *domol (slow), latent (latent), and exigibil (exigible)*. Despite two of these items being cognates, due to their low-frequency, they were known by fewer than 30% of participants.

Ideally, to be able to conclude that all items in a test are of reasonable difficulty, and therefore, are able to discriminate well among students with different abilities, no items should elicit a perfect score. In the case of vocabulary tests, however, as items are chosen
from different frequency bands, this criterion becomes somewhat problematic, as one would expect lexically able students to have mastered the 2k words. The RomVLT seems to perform acceptably regarding this criterion, however, the four items on the VLT, which are above the 2k band, should be revisited. Similarly, items that are too difficult will elicit very few, erratic responses, which means again, they cannot differentiate well amongst students. The Romanian test appears to be abiding by this criterion as very few items have a hit rate below .30 and all of them are from the 10k frequency level. However, the two words from the UWL described above should be reconsidered prior to further studies in the case of Romance language speakers.

C. Lexical proficiency and cognate recognition

The last issue regarding RQ2 that should be addressed is whether proficiency influences cognate recognition. Studies reported in Chapter 2 suggest that in some cases low-proficiency students might not be as perceptive to inter-linguistically congruent items as their more proficient peers due to limited metalinguistic awareness or cumulative LL experience. To investigate this, I split the results into high and low scores, using the median split as this provides two groups equal in size.

<table>
<thead>
<tr>
<th></th>
<th>Group</th>
<th>N</th>
<th>Total</th>
<th>Cog Total</th>
<th>Cog 2k</th>
<th>Cog 3k</th>
<th>Cog 5k</th>
<th>UWL Cog 10k</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>M</td>
<td>20</td>
<td>73.60</td>
<td>42.15</td>
<td>9.60</td>
<td>10</td>
<td>4.50</td>
<td>13.85</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
<td>7.75</td>
<td>3.13</td>
<td>0.75</td>
<td>0</td>
<td>0.69</td>
<td>1.69</td>
</tr>
<tr>
<td>Low</td>
<td>M</td>
<td>20</td>
<td>45.25</td>
<td>29.30</td>
<td>8.00</td>
<td>8.05</td>
<td>2.45</td>
<td>9.20</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
<td>9.39</td>
<td>4.99</td>
<td>1.17</td>
<td>1.76</td>
<td>0.89</td>
<td>1.79</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>40</td>
<td>59.42</td>
<td>35.73</td>
<td>8.80</td>
<td>9.03</td>
<td>3.48</td>
<td>11.53</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
<td>16.68</td>
<td>7.70</td>
<td>1.26</td>
<td>1.58</td>
<td>1.30</td>
<td>2.92</td>
</tr>
</tbody>
</table>

Table 4.9 reports the means and standard deviations for the total scores, the total cognate scores, and cognate scores on the different levels based on the median split into High and Low performers on the VLT. First, it is noticeable that the test is capable of distinguishing between high and low performers as indicated by the total mean scores. The
High group \((M = 73.60, SD = 7.75)\) did not only outperform the Low group \((M = 45.25, SD = 9.39)\) but performed more consistently as well, as the standard deviation scores suggest. Based on the Shapiro-Wilk test, normality can be assumed in both groups (High: \(W = 0.96, p = .59\); Low: \(W = 0.96, p = .48\)) and the Levene’s test indicated that the assumption of homogeneity of variance between the groups is also tenable \((F(1, 38) = 0.07, p = .79)\). Hence a t-test was performed to explore whether the difference between the two groups is significant. The two-sample t-test confirmed \((t(2, 38) = 10.42, p < .001)\) that the difference between the two groups is significant with a large effect size \(R^2 = .86\).

The next step in the analysis was to check whether the High group recognised more cognates than the less proficient group. The means and standard deviations suggest that the High group \((M = 42.15, SD = 3.13)\) recognised more cognates than the Low group \((M = 29.30, SD = 4.99)\) and their scores also show more consistency. As the Low group’s distribution violates the normality assumption, the Wilcoxon rank-sum test was employed to examine the difference. This indicates that lexically more apt students recognise significantly more cognates than the Low group \((W = 397.5, p < .001)\), with a large effect size \(R^2 = .84\).

A mixed model was fitted to the data then to further explore the relationship between proficiency and cognate recognition. This was a by-item analysis. Compared to the null model, introducing a random intercept for items \(X^2(1) = 12.21, p < .001\) and a random slope for proficiency significantly improved the model \(X^2(2) = 28.76, p < .001\). These two models confirm that scores vary across items and proficiency, which is explained by the significant change in the maximum likelihood ratios. Next, the cognate status \(X^2(1) = 5.41, p < .05\) and proficiency \(X^2(1) = 113.29, p < .001\) are added as fixed effects. The two fixed effects, resulted in a significant improvement in the model. With the interaction, the model still improved and this provided the final model, \(X^2(1) = 6.04, p = .01\). The effect size values are as follows: \(R^2_M = .31\) and \(R^2_C = .94\), which means that the model explains at least 31% of the
variance, but this is significantly improved by accounting for the random effects.

There is no indication of multicollinearity as all values of tolerance > .2 and VIF values < 5. The main effects of proficiency ($b = -12.93$, $t(88) = -2.68$, $p < .001$) and cognate status ($b = -26.77$, $t(88) = -9.56$, $p < .001$) indicate that if all other variables are ignored, the low-proficiency group is significantly different from the high-proficiency group, and as the negative $b$ value suggests, the Low group attained lower scores. Similarly, compared to cognates, non-cognate scores are negatively associated, indicating that cognate scores are predicted to be higher than non-cognate scores. However, since the interaction between proficiency and cognate scores, as a higher order interaction, is significant ($b = -2.65$, $t(38) = -2.22$, $p = .03$), the main effects should be disregarded. The following plot facilitates the interpretation of the interaction.

![Figure 4.5 Interaction between proficiency and cognateness on the VLT](image)

**Figure 4.5 Interaction between proficiency and cognateness on the VLT**

The significant interaction suggests that cognate items facilitate more hits for both groups. However, in the case of the lower proficiency group, the gap between cognates and non-cognates becomes more accentuated. Specifically, this means that the facilitative effect
of cognates has a larger effect in the case of the lower group.

Subsequently, the same relationship between lexical proficiency and cognates will be explored on the Romanian test. Prior to this, however, it should be clarified that instead of using the High-Low binary variable from the VLT, I have decided that the median-split will be based on the Romanian results in this case. The reason for this is twofold. On the one hand, it is of interest how Romanian lexical proficiency influences cognate recognition in Romanian, instead of the effect of English proficiency on Romanian cognate recognition. On the other hand, based on the strong L2-L3 correlation between the overall scores reported earlier and a comparison between the two categorical variables, which confirms 80% similarity (i.e. 32 out of 40 students were assigned to the same category [High or Low] on the two tests), one can be confident that the difference in the two splits is minuscule.

Table 4.10 Cognate results on different frequency levels on the RomVLT

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Total</th>
<th>Cog Total</th>
<th>Cog 2k</th>
<th>Cog 3k</th>
<th>Cog 5k</th>
<th>Cog 6k</th>
<th>Cog 10k</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>20</td>
<td>75.70</td>
<td>39.50</td>
<td>9.35</td>
<td>8.40</td>
<td>4.20</td>
<td>13.65</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.11</td>
<td>5.07</td>
<td>0.59</td>
<td>1.76</td>
<td>0.62</td>
<td>2.58</td>
<td>1.62</td>
</tr>
<tr>
<td>Low</td>
<td>20</td>
<td>38.70</td>
<td>20.75</td>
<td>7.70</td>
<td>4.60</td>
<td>3.20</td>
<td>4.30</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.00</td>
<td>6.41</td>
<td>1.53</td>
<td>2.09</td>
<td>1.24</td>
<td>2.25</td>
<td>1.39</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>57.20</td>
<td>30.12</td>
<td>8.53</td>
<td>6.50</td>
<td>3.70</td>
<td>8.97</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21.49</td>
<td>11.08</td>
<td>1.41</td>
<td>2.71</td>
<td>1.09</td>
<td>5.30</td>
<td>2.11</td>
</tr>
</tbody>
</table>

In comparison to the VLT, Table 4.10 also confirms that the Romanian test can distinguish High and Low performers well. The results show evidence towards the High group ($M = 75.70, SD = 9.11$) performing better and more consistently than the Low group ($M = 38.70, SD = 12$). The assumption for normal distributions was tenable for both variables (High: $W = 0.97, p = .85$; Low: $W = 0.92, p = .09$) and homogeneity of variance is confirmed as well ($F(1, 38) = 2.77, p = .10$). Thus, to check the difference and its magnitude, a $t$-test was applied. This reveals that the difference is significant ($t(2, 38) = 10.98, p < .001$) with a large effect size $R^2 = .87$.

Extrapolating from the cognate means and standard deviations, it can be noticed that
the High group ($M = 39.50, SD = 5.07$) recognised more cognates and more consistently so than the Low group ($M = 20.75, SD = 6.41$). Since the assumptions for normality (High: $W = 0.97, p = .68$; Low: $W = 0.94, p = .18$) and homogeneity of variance ($F(1, 38) = 1.87, p = .18$) are tenable using the usual tests, a two-sample *t*-test was conducted to explore the differences. This confirms that there is a significant difference in cognate recognition between the High and Low groups ($t(2, 38) = 10.26, p < .001$) with a large effect size $R^2 = .86$. It is also observable that on the 6k level, the High group had a mean score of 13 out of 17, whereas the Low group only reached a mean just over 4. This difference lends way to speculation that low-proficiency students struggle with infrequent cognates.

Additionally, a mixed-model analysis was conducted to explore the relationship between proficiency and the cognate status of items on the RomVLT in a by-item analysis. Compared to the baseline model, introducing a random effect for proficiency and items results in an improved model, $\chi^2(3) = 34.68, p < .001$. This confirms that there is significant variation between items and proficiency within the model. Next, I introduced cognateness as a fixed effect, followed by proficiency. Cognateness, did not yield a significant improvement to the model, which means that on average, items on the RomVLT were recognised to the same extent regardless of cognacy. By contrast, however, proficiency improved the model significantly $\chi^2(1) = 137.54, p < .001$. The interaction between the two fixed effects was not significant either, meaning that cognate recognition does not vary across proficiency. The final model thus explains at least 42% of variance as $R^2_M = .42$ and $R^2_C = .94$. Multicollinearity does not affect the model as both the tolerance (> .2) and VIF values (< 5) are within the expected range. By observing the main effects, it can be concluded that cognateness does not influence item recognition ($b = 4.21, t(88) = 1.11, p > .05$). The fixed effect for proficiency ($b = -41.11, t(89) = -17.87, p < .001$) again, confirms that there is a significant difference between the number of items recognised by the High and Low groups.
Figure 4.6 demonstrates that the RomVLT clearly distinguished between students’ of low and high lexical dexterity, however, the cognateness of words does not affect the model. In comparison to the English test, where cognates seemed to provide a clear advantage for both groups, this difference suggests that the tests differ qualitatively and this might be an artefact of including academic words in the VLT. In the interest of exploring this further, the ensuing graphs will present the percentage of correct answers on the different frequency divided by proficiency.

Based on the percentage of hits on the different frequency levels on the VLT, it can be noticed on Figure 4.7 that on the first two levels, lexically able students recognise the same amount of cognate and non-cognate items, whereas the low-proficiency group can capitalise on cognates affordances as non-cognate items were recognised by fewer students. Furthermore, the graph illustrates well that difference between cognate and non-cognate items is more accentuated in the case of the low-proficiency group.
Figure 4.7 Interaction between proficiency and cognateness based on frequency on the VLT

Figure 4.8 Interaction between proficiency and cognateness based on frequency on the RomVLT

With regard to the RomVLT, Figure 4.8 illustrates that while the high-proficiency group performed comparably in terms of hits to cognate and non-cognate items, the Low
Results

group benefitted from the cognate facilitation effect. Based on the aforementioned results, there seems to be substantial evidence to suggest that cognate recognition is proficiency dependent, and students with lower overall lexical knowledge are prone to miss (or not attempt) more items, regardless of the affordances that cognates potentially provide.

In sum, the data analysis carried out in order to address RQ2 revealed that there is a strong relationship between the L2 and L3 lexical knowledge proven by the tests employed and the form-meaning mapping skill tested. This is confirmed by the robust correlation between the total scores and the regression models. The interrelatedness between the two lexicons is even more emphasised in the case of cognates, insofar as, cognates in one language alone explained more variance in the total score of the other language than non-cognate knowledge. Based on the analyses, there is also significant evidence that English cognates are more easily recognised, and provide more affordances than Romanian cognates compared to non-cognates, albeit not all cognates have been recognised in either of the languages. This could be purportedly ascribed to the fact that Romanian is an interlanguage as well and, as the tests show, lexical proficiency in the L2 and L3 are very close to each other. Moreover, as argued before, this might be an artefact of the academic words featured on the VLT. Considering only the cognate scores, the high-proficiency group outperformed the low-proficiency group, and this difference pertains to both languages tested. Importantly, the mixed models highlighted that there is a significant interaction between proficiency and the cognate status of the items, and that, with regard to the English test, low proficiency learners can benefit from the presence of cognates on the test. This, however, does not stand for Romanian items. Finally, in section 4.8.3, some problematic items have been encountered and in most cases the reason for this seemed to be cognateness, especially in the case of the VLT. This, in turn, confirms the initial hypothesis that item selection with respect to cross-linguistic influence can pose potential problems to testing.
4.8.4 Exploring the relationship between test scores and learner-related factors

This subsection aims to bring the empirical results on the tests and the questionnaire (Appendix 3) data closer in order to aid a triangulation of specific aspects, add to the validity of the tests, and reveal more information about issues focal to multilingualism and hence, the current thesis. Critically speaking, the questionnaire is not a stand-alone study, due to the limited amount testing time, and therefore, the number of questions. Thus, its main objective is to complement the test results and shed further light in an exploratory fashion on certain selected questions that were of interest. Accordingly, the questionnaire comprised eight questions that can be categorised as three five-point Likert-scale questions, three multiple-choice questions, one question about the age they started learning the L2 and L3, and one open question that enabled participants in the SG to comment on the EI. The results were compared using the most appropriate Spearman rho rank-order correlations.

The first question addressed the age of onset of learning Romanian and English. Questions 2 and 3 asked students to rate their confidence in using Romanian in an informal (everyday) and a formal (job interview/presentation) setting. Since Molnar’s (2010) study did not find any correlation between the age when participants started learning the language and their language proficiency, there was no clear expectation regarding this question. However, as the results reveal, there is a moderately significant negative correlation between the age participants started to learn Romanian and both their confidence and proficiency ratings \( (\rho = -0.53 \text{ and } -0.66; \ p < .001 \text{ and } N = 44 \text{ in all cases within this section}) \), and RomVLT scores \( (\rho = -0.49) \). Meaning that the younger students started learning Romanian, the more confident and proficient they feel in using it and this is supported by the RomVLT scores as well. There was also a negative association between the age of onset of learning English and VLT scores \( (\rho = -0.39) \). However, no significant correlations emerged between the VLT scores and English confidence ratings or age of onset of learning English.

Table 4.11 shows students’ ratings from 1-5 (1 being not confident and 5 being very...
confident) on Questions 2 and 3. This indicates that seven participants (27%) from the SG would not feel confident using Romanian in everyday situations, whereas only two participants (11%) from the CG chose the same answers.

**Table 4.11 Summary of answers to Questions 2 and 3**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Question 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Question 3</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Study</td>
<td>5</td>
<td>2</td>
<td>12</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>n = 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, 13 participants (50%) in the SG, versus seven (38%) in the CG, would not feel confident using Romanian in a formal setting. Considering these results and comparing it to Table 4.1, it can be asserted that students’ self-ratings more or less represent their lexical knowledge as demonstrated by the RomVLT, which provides further evidence for the validity of the test.

**Table 4.12 Summary of answers to Questions 4 and 5**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Question 4</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Question 5</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Study</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>n = 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Control| 1 | 4 | 4 | 5 | 4 | 0 | 3 | 7 | 5 | 3 |
| n = 18 |   |   |   |   |   |   |   |   |   |   |

Table 4.12 demonstrates that 16 participants (61%) in the SG (versus nine in the CG) would feel more confident using English in a workplace situation (Question 4), despite the fact that they live in Romania and have been exposed to their L2 for longer. The answers to Question 5 indicate that only five (11%) participants overall would not feel confident in using English in the target-country, while the rest suggest they would be fairly to very confident (89%).

Through students’ self-reflection, Question 6 had the aim of understanding participants’ thinking process when using English. The answers reveal that 16 participants
in the SG and 24 in the CG (91% in total) ‘think of suitable words in Hungarian (L1) and translate them’ rather than capitalising on their L2. This is chiefly a result of students relying on their L1 while learning English. This would support the notion of EI on cognates being beneficial for students. However, the findings suggest that the underlying process of cognate recognition might be intuitive regardless of whether students are aware of this or not. This is elaborated on in Chapter 6.

The last two questions allowed students to reflect on their opinion on their L2 and its use for learning English (Question 7), and to comment on the EI, in the case of the SG (Question 8). The responses reveal that 25 participants (56%) think that their Romanian knowledge ‘definitely helps them’ or ‘in many cases it comes in useful’ in learning or functioning in English and only five students (11%) think that ‘it makes no difference’ or ‘it can be confusing’. This points in the direction that students do find their L2 beneficial overall and have a positive predisposition towards it.

The comments given by students provided insights on the value of the cognate instruction and their perception of the link between Romanian and English. In essence, most students considered the instruction useful and pointed out the benefits of making connections between the two languages. Nevertheless, some students commented that due to their limited Romanian, the instruction or further instructions would not be helpful. Interestingly, one student reported that the instruction did not help, as if he/she does not already know the word in one language, the meaning would not be known in the other language either. This suggests that it is possible the search for meanings is automatic in the two languages, at least for this particular student. Students’ comments reflect these issues:

a. I found this handout useful and helpful in making connections between cognates. My opinion is that it is and it would be helpful to teach and learn a foreign language with the help of exercises and handouts on cognates. […] (Stud3008S)

b. It only helped because I was trying to make a connection intentionally. I would not necessarily do that. I'm not used to it. It would have helped when we started
to learn English. But it could have been useful in learning Romanian. If our knowledge of both languages would be on the same level, we could benefit from it. (Stud3010S)

c. The handout helped making a connection between English and Romanian. It would be easier to learn if we received similar instructions. (Stud1042S)

This section has demonstrated the importance of gathering and analysing not just participant-related data, such as age of acquisition, but information on their perception of the background languages, proficiency and confidence ratings. It appears that these information complement the results elicited by the tests and allow a more finely tuned interpretation of the study. As such, it is important that according to students, their L1-L3 link predetermines their use and processing of English without, at least conscientiously, relying on the L2. If in the case of adult and proficient learners the Hungarian-English link is so enshrined within their FL competencies, there might be clear value in advocating lexical interconnections at earlier stages of language development.

The study presented in this chapter tapped into several important lexical properties that play a role in the conceptualisation of the VS of multilinguals and through this, it also raised some questions of prevalence. The following chapter revisits these findings and aims to add a finer level of granularity to these issues and the relevance they bring to the field of lexical studies.
Chapter 5: Study Two (S2)

S1 revealed some fundamental issues concerning lexical assessments in multiple languages and the strong relationship between a Romance L2 and English as an L3. Importantly, several problematic issues have been pointed out that pertain to languages being inextricably connected within the ML. This is due to CLIs between the examined languages and has been scarcely addressed regarding VS. This study, therefore, builds on, deliberates, and scrutinises these findings and, as the forthcoming sections reveal, takes it a step further in order to gain deeper insights into this interconnection that, as it appears, indisputably affects word recognition.

Chapter 3 has already emphasised some limitations of the VLT and the empirical findings above further support this in relation to the language combination and proficiency levels tested. Specifically, it appears that on the VLT, the first two bands cannot differentiate between students reliably and the inclusion of academic words has a negative impact on the interpretation of the results. In conjunction with this, there is substantial evidence for item facility indices being distorted by cross-linguistic similarities. Furthermore, the findings corroborate that cognates do represent a special bridge between the languages and, since apart from frequency, the cognate status of items also affects lexical testing, further investigation into this phenomena seems justified. Based on Chapter 2, it has been concluded that lexical processing and use is influenced by cognateness, and this, in combination with the results in S1, lends support to the belief that the special status of cognate words within the ML deserves further investigation.

In this attempt, primarily two important aspects have been reconsidered for the study discussed herein compared to S1. On the one hand, the test format was kept more or less the same, however, instead of testing only samples from five frequency bands, using the VST allows testing words from 14 different levels. Moreover, the proportion of correct answers and distractors is reduced within the VST to four in each cluster. This, on the other hand,
enables investigating the special status of cognates within the lexicon through RTs (see Chapter 3). In light of these differences between the two studies, the ensuing sections provide additional background for the present study by making explicit connections between the two studies, and offer a detailed account of the methodology and data analyses. Finally, the results obtained are systematically addressed in conjunction with the formulated RQs. The findings of the two empirical studies are compared and further discussed in Chapter 6.

5.1 Objectives and Research Questions

The present study has set out with two primary objectives in mind. The first objective was to attempt replicating the findings in S1. Similar results would point in the direction that the evidence presented so far is not simply an artefact of the methodology, tests, or analyses employed, but would also lend further credit to the conclusions and the validity of the arguments. It is felt that the intricacies of vocabulary retrieval, and how this is affected by CLIs, can be further disentangled by examining speed of lexical access, even if in this case this is a relative measure that pertains to the multiple-choice test formats and the methods employed. This is, then, the second objective.

Consequently, as S1 produced contradictory findings to Nagy et al. (1993), for example, to the extent that while they seem to support the facilitative effect of an awareness-raising intervention, conclusions herein (and statistically speaking Molnar’s [2010] as well) indicate that despite students’ comments, the instruction on cognates had no effect on VS. Therefore, the data seems to corroborate Hall (2002), for instance, and echo the view that cognate recognition is indeed an intrinsic phenomenon with regard to the languages, age, and proficiency of participants. In the present study, I considered important to retain this aspect of the research and, consequently, RQ1 remained unchanged. The reasoning behind this is that the number of cognates in the VLT, especially the academic words, retrospectively, seem to have partially exacerbated the validity of the results. Therefore, it appeared logical that prior to giving full support to the fact that proficient learners rely on
such affordances automatically, additional evidence was needed to strengthen this claim.

Corresponding to this, the previous study presented strong evidence towards the close L2-L3 relationship. Proficiency, in terms of lexical dexterity, in one FL proved to be an indicator of proficiency in the other, converging with results put forward by Molnar (2010) and Muñoz (2000). Considering the implications of this, from pedagogical and testing points of view alike, it was felt that confirming the previous results by using a different lexical test, and its Romanian and Hungarian versions, would further evince and perhaps generalise the arguments asserted in this thesis. To this end, RQ2 was maintained as well.

S2 also involves a temporal measure to gain additional insights into the cognate facilitation effect and CLIs. This became RQ3 and, aligning with S1, the last part of the data analysis triangulates the test results obtained and the questionnaire data. Thus, the RQs for the current study are as follows:

- **RQ1:** What is the effect of explicit instruction on English-Romanian cognates on the vocabulary knowledge of learners of English?

- **RQ2:** What is the relationship between overall lexical knowledge and cognate recognition in the L1, L2, and L3?

- **RQ3:** What is the relationship between lexical knowledge and speed of lexical access?

Generally, the hypotheses regarding these RQs converge with the ones formulated in the preceding chapter. However, as this is a separate study, it seems appropriate to briefly reiterate these by centralising incongruous factors in comparison to S1.

One focal aspect of this thesis is potential cognate affordances in terms of lexical recognition and the effect these, therefore, have on language testing. RQ1 explores whether an EI on cognates enables participants in the experimental group to substantially capitalise on such affordances and thus, recognise more cognates than students in the CG. Previous studies, involving various language constellations and scripts (e.g. Hancin-Bhatt and Nagy,
1994; Nagy et al., 1993; Otwinowska, 2016) extrapolated that relying on cognates facilitates reading and listening, and perhaps can serve as a motivational factor for students. S1 has indicated that more cognates are recognised than non-cognates, especially in English and by participants with lower proficiency. However, there was no indication that the experimental paradigm had an effect. Nevertheless, the fact that not all cognate items were recognised by all participants and there were individual variations between participants, proficiency and frequency levels, suggests that cognate recognition is not ubiquitous across the board and there might be room for improvement. Consequently, based on the above and the typological distance between Romanian and English, it is hypothesised here that Romance language speakers may benefit from EI.

Building on S1, RQ2 addresses the L2-L3 relationship from various perspectives. The analyses in the previous chapter and to some extent Molnar’s (2010) findings imply a strong relationship between Romanian and English lexical knowledge. These obtain that more proficient students will also demonstrate increased knowledge in the other language, which, notably due to CLIs between the tested languages is most pertinent on the level of cognates. Furthermore, the cognate facilitation effect is also investigated through an item analysis. In other words, I explore how certain items on the tests are influenced by cognate knowledge and if there is a cognate facilitation effect, how this can be addressed from a testing perspective. By exploring this relationship between the languages enables examining a variety of factors of importance affecting vocabulary tests that have been neglected in the past.

Since S1 showed that languages are intertwined within the lexicon and cognates indeed represent special items, of interest was whether this effect is evident in a temporal measure as well. Accounting for RTs in the design had mainly three reasons. First, due to the limited number of vocabulary studies that investigate the relationship between VS and RT, I became interested in whether the speed of lexical access is in line with VS in the three different
languages. In other words, is the VST with a temporal measure capable of differentiating between proficiency in different languages? Evidently, Hungarian NSs should have the highest and most consistent scores on the HUVST and this should reflect in their RT as well. On the other hand, are more proficient learners likely to display faster lexical access than their peers with lower proficiency? Lastly, in the case of multilinguals, is it possible to notice a frequency effect in all three languages or only in the L2 and/or L3, or none?

Second, RT measures will allow to compare the results to previous findings, which state that cognates have a processing advantage in comparison to non-cognates or pseudo-words, thus, they are recognised and translated faster, and more accurately (e.g. De Groot & Keijzer, 2000). In this case, this means that form-meaning mapping should be faster for cognates than non-cognates. It also seems fair to hypothesise that multilateral cognates are to be recognised faster than bilateral cognates and, respectively, faster than non-cognates. Additionally, it is expected that cognates will have a facilitative effect in the case of the L2 and L3 tests, albeit not in L1.

The development and use of a FL are influenced by a number of external factors, such as students’ perception of the link between their background languages, predisposition towards the state or TL, and (psycho)typology. Therefore, the questionnaire aims to take our understanding of influential factors beyond the interpretation of test scores and the effect of cognates, and shed light on the issues that affect multilingualism, however, cannot be effectively captured in a vocabulary test. To accomplish this, the questionnaire collected data on participants’ age-of-acquisition ratings, self-rated proficiency, and their perception of L2-L3 link. These factors are considered fundamental to LL and assessment, and as S1 has demonstrated, the questionnaire data does not only complement the vocabulary test results but aids the explication of underlying issues that otherwise would go unnoticed.

With these goals in mind, the next sections elaborate on the design of S2, ethical considerations, participants, instruments, data collection, and the analyses of the results. A
key element of these introductory sections is to highlight essential points in which S1 and S2 diverge, the benefits of these methodological decisions, and the rationale behind them.

**5.2 Design of Study Two**

This section will explicitly focus on divergent aspects of design in comparison to S1 (see section 4.2). Chapter 3 indicated that instead of the VLT, S2 opted for the VST (ENVST) and its Romanian version, the ROVST. Moreover, while S1 was a pen and paper test, it was decided that S2 will be an online experiment (in both the psycholinguistic real-time and web-based senses) in order to accommodate a temporal measure. In conjunction with this, to compare RT measures to a baseline, i.e. the L1, the Hungarian version of the VST (HUVST) was also included in the test battery. All three tests, the questionnaire, and consent forms have been uploaded to Qualtrics.com, which is a survey website that records the number of clicks, time passed between clicks, and total submission time for each question. At the time of design, automatically uploading full sets of questions to the website was impossible, so each question had to be inserted individually with the appropriate time-measure. Furthermore, although setting up the scoring for the questions was possible at the design stage, after trial sessions and all participants completed the test battery, it transpired the website has a technical glitch causing jumbled answers within items. This meant that at the data analysis stage an answer sheet (key) had to be created in MS Excel and applied to the answers to obtain the correct scores.

Additionally, as noted in S1, once students reach the lowest frequency level their answers (including hits, but mainly attempts) become infrequent. This suggested that as students feel unsure of the answers, they potentially give up and finish the test without even considering these items. To rectify, it seemed justified to set up all items as obligatory (forced-answer) questions, which also increased the length of experiment.

It should also be mentioned that due to the way the survey website works, and the lack of touch-screen monitors, the survey was completed using a standard two-buttoned mouse.
This, nonetheless, can be surmised a more natural method from a student perspective, as no time was actually required to train learners in using specially assigned buttons on a keyboard (or other specialised equipment) – using a standard mouse is a process that participants are accustomed to. Although this method is frequently used by language teachers in lack of a laboratory (e.g. Laufer and Nation, 2001), it needs to be acknowledged that this elicits less precise measures compared to E-prime. However, it is important that currently no RT measure captures accurately the duration of lexical accessibility in human brains (Tanabe, 2016).

Additionally, an RT measure becomes relatively high considering the four definitions and a non-contextual sentence. Therefore, these RT measures should be taken as relative under the given conditions, which, concurrently overcomes the possibility of confusion over the four answer and a further ‘pass’ button on a regular keyboard reported by Tanabe (2016). Strictly speaking, highlighting an item with a standard mouse and pressing a submit button provides confidence that subjects submitted exactly the answer they intended to and did not get confused over five buttons that automatically submit the answer without allowing participants to review the choice (e.g. Tanabe, 2016). In turn, considering these aspects of the procedure, the study becomes easily replicable to any practitioners or researchers as only a mouse and computer (or phone etc.) with internet access is required.

The tests were ordered and administered according to proficiency: English, Romanian, and Hungarian, and without counterbalancing languages or items across frequency levels. The argument behind this, additionally to the ones mentioned in section 4.2, is that counterbalancing items between languages would potentially add switching delays to the temporal measure, thus lengthening the study and increasing the cognitive load on students unnecessarily. It was also decided that splitting the questionnaire into two parts would benefit the students in terms of task demands. Therefore, questions related to the age when they started learning different languages, and self-rated confidence and proficiency levels
Ethical Considerations

In conjunction with S1, the Open University’s guidelines were followed (see section 4.3 for these). Ethical approval for S2 was granted by the Human Research Ethics Committee of the OU (HREC/2015/1626/Szabo/2).

5.3 Ethical Considerations

In order to strengthen and bring further evidence for the findings in the previous chapter, I felt that it is important to maintain the same DLC and proficiency levels in the present study as well. Thus, participants belonged to Partium Christian University. Initially, the aim was to test Hungarian speakers enrolled for the same degree at the state university, the University of Oradea, however, due to the difference in examination periods and conferences this was unachievable (see Chapter 7 for limitations). Nevertheless, the research involved 54 Hungarian NSs, who live in Romania and therefore are exposed to Romanian as an L2, and study for a BA in English Language and Literature (first-year \( n = 23 \), second-year \( n = 17 \), third-year \( n = 14 \)).

All participants considered themselves NSs of Hungarian on the questionnaire. Only three students considered themselves as NS of Romanian, while the rest of them started learning Romanian around the age of 7 \( (N = 54, \min = 0, \max = 8, SD = 2.07) \). Regarding English, the mean age of acquisition was 9 \( (\min = 2, \max = 14, SD = 4.4) \). Furthermore, the background questionnaire reveals that 76% of participants never learnt French and those who did, started on average at the age of 15 \( (\min = 11, \max = 21) \) and very few learnt Spanish (four) or Italian (two) respectively as an L4. Just over 50% of students learnt German as an L4, some of them as a beginner and some as advanced. However, as the current studies focus...
Chapter 5: Study Two (S2)

on the effect of bilingualism and the relationship between an advanced L2 Romance language and English L3 (DLC), students’ German (or other non-dominant Ln) VS was not tested. Similarly to S1, the number of participants in the SG and CG ($n = 22$ and 32) became unevenly distributed due to lack of attendance on the day of the data collection. No students had scores below 16 (cut-off score used in S1) on either of the tests. In fact, on the ROVST 25 was the minimum score, and thus no participant data has been discarded prior to the data analysis.

5.5 Instruments

To briefly recapitulate the reasons for opting for the multiple-choice form-meaning mapping vocabulary test format designed by Nation and Beglar (2007) instead of the multiple-matching format featured on the VLT, some internal and external factors should be mentioned. External factors include the fact that the VLT by design is a diagnostic tool composed of sampled items from four frequency levels and the UWL. Following S1, it was learnt that these four levels make it problematic to estimate VS in English and with the focus group of these studies, interpreting the scores elicited by the academic words that are in majority also cognates becomes somewhat problematic. Therefore, to maintain the focal aspect of vocabulary knowledge within the studies, that is form-meaning recognition, the VST appeared to be well-suited, for it comprises randomly sampled items from the first 14k most frequent words of English. Thus, estimating VS from the test scores becomes more meaningful, there are no academic words to be espoused as a compilation criteria, and the prospects of reduced bias make the interpretations more sanguine.

The internal factors are those that pertain to the purposes of this thesis. In comparison to the VLT, the format of the VST features one word at a time and four definitions to select the correct answers from. Consequently, measuring RT becomes more straightforward for interpretation. Additionally, on the VLT a number should be provided for the correct answer, which would probably also involve students’ typing skills and thus result in more individual
differences manifested in probable response latencies. The VST’s format was compatible with the question-types offered by Qualtrics.com and offered a familiar layout to all participants.

S1 served as evidence for the fact that the Questionnaire complemented well the lexical scores elicited by the tests and aided some quintessential reflections and conclusions in relation to the multilingual aspect of the present thesis, including age of acquisition, self-rated proficiency and confidence levels, and importantly, students’ perceived distance between their languages. The logical entailment of this was that the test-battery described above was completed by a Questionnaire in this study as well. These instruments are presented in the Appendices.

5.6 Data Collection

In line with the procedure set out in section 4.6, the data collection for the present study comprised three stages. First, participants were divided into the experimental conditions, following the random groupings they use for the seminars. The SG received EI on cognates in the form of a handout (Appendix 4) that points out English-Romanian structural similarities. Additionally, in this study a cognate recognition task (Appendix 8), in which students had to underline cognates in three short and identical paragraphs in Romanian, English, and Spanish, was also added to this stage. The assumption here was that pointing out structural similarities and providing examples is perhaps not as motivating or challenging as reading an abstract from the movie Madagascar in three parallel languages and identifying cognates. As this task only served the aim of raising awareness of cognates in a short time-frame and cannot be regarded as a pre-test for cognate-recognition strategy, its results are not included in any of the analyses. Students, in actuality, were so motivated by this task that they asked in several different groups to keep it for future pedagogical purposes as reference.

The next stage featured the VSTs, ensued by the post-test questionnaire. The
completion-time slightly increased from 40 to around 50 minutes due to the addition of the Hungarian test and some extra questions in the questionnaire.

This study was also conducted during normal teaching hours in a phonetics lab, which is utilised by participants on a weekly basis and all computers are identical in terms of speed, screen size, and keyboard/mouse type. To discourage students using additional resources, such as dictionaries, phones, and switching between browsers, I made them aware that the time is limited, however, it was not mentioned that all items and answers are timed individually, to avoid rushing and potential guessing as a consequence. Moreover, during the experiment, I was present in order to explain the aims of the study, rectify technical issues in the case they arose, answer questions, and ensure that students work individually without cheating. No technical issues were recorded throughout the data collection. None of the participants reported any visual impairment or physical difficulties with operating a standard desktop PC.

## 5.7 Data Analysis

Once the recorded data was downloaded from Qualtrics.com, any unnecessary information, such as IP addresses, data and time stamps, had been removed from the dataset. As has been mentioned previously, due to a website error, the automatic scoring has not worked, which meant that a digital scoring sheet had to be created in MS Excel and applied in order to derive a binomial dataset comprising 1 for correct answers and 0 for incorrect ones. To prepare the data set for descriptive and inferential analyses, it was uploaded to R (R Core Team, 2015), which contains the same packages as in S1.

The data analyses hitherto described and employed in Chapter 4 will be strictly adhered to in the present chapter as well. In addition, RT measures were analysed using multilevel regression analysis. Finally, as section 3.7 on the instruments explained, apart from cognateness and frequency, the Romanian test for S2 was designed to account for POS and word length, operationalised as the number of syllables. Therefore, in order to find the
most parsimonious model that explains the results, similarly to Willis and Ohashi (2012), inter alia, multiple linear regression models were employed to explore the effect of these variables on accuracy and RT. In this case, this means that a stepwise variable selection procedure was applied (identical to the procedure employed in S1), which entails iteratively adding or removing predicting variables to find the model that best explains the data and the interaction between the outcome (or dependent) and covariate (independent or explanatory; predictor) variables (see Field et al., 2012). To ensure the assumptions of normal distributions and homogeneity of variance are met, the previously used Shapiro-Wilk and Levene’s tests were applied.

5.8 Results

The previous sections focused on the RQs and methodology that are rooted in, incrementally build on, and improve on S1. Now, the focus is turned to the presentation of the results and data analyses pertaining to these questions. In conjunction with S1, however, an investigation into the reliability and some aspects of validity are offered in the first instance.

5.8.1 Issues of reliability and validity

The findings in S1 seem to imply that the test results obtained indicate an acceptable level of reliability and stand, to some extent, against the scrutiny of validation. By acknowledging the importance of these aspects of reliability and validity, in order to avoid repetition and ensure cohesion, I strictly focus on the criteria set out in section 3.9.

The item reliability index for the ENVST was acceptable, using Cronbach’s alpha, at $\alpha = .83$, with no items significantly improving or worsening the score if dropped. The internal consistency of the test based on the 14 levels is also acceptable at $\alpha = .80$ (lower and upper $\alpha = .70$ and .91) and dropping any of the levels would only decrease the alpha coefficient.

Concurrently, Cronbach’s $\alpha = .84$ for ROVST items, with all items contributing in the +/- .01 range if the item is dropped. The reliability on the 14 levels is also acceptable at $\alpha = \ldots$
.83 (lower and upper $\alpha = .74$ and .92). Due to a ceiling effect on the HUVST, which was expected as participants were Hungarian NSs, the item-reliability scores emerged at $\alpha = .65$ (lower and upper $\alpha = .50$ and .79) and for the 14 frequency levels at $\alpha = .56$.

These lower reliability figures indicate that the variance of items and correlations between items are reliable on the ENVST and ROVST, however, this is not the case on the HUVST. Due the ceiling effect on the HUVST, lack of variance, and the interest in exploring the L2-L3 relationship, only the results in the L2 and L3 will be compared, except in the case of RQ3 where RTs are compared rather than vocabulary scores.

In comparison to S1, lower reliability scores can be observed in the case of L2 and L3 tests as well. Arguably, this is largely due to the fact that instead of 90 items as on the VLT and RomVLT, the number of items on these tests has been reduced to 70 (from 140 as on the original VST). Nevertheless, based on the acceptance level of $\alpha > .80$ (e.g. Weir, 2005), both the ENVST and ROVST seem to have elicited reliable scores with a significant level of internal consistency. Considering only the Likert-scale questions on the questionnaire ($N = 54$, no. of questions = 13), Cronbach’s $\alpha = .87$, which indicates high internal consistency.

Next, the item statistics are interpreted. Inter-item correlations, or item homogeneity, indicate that the average correlation for the ENVST is .07, for the ROVST is .09, and for the HUVST is .06. These figures indicate that the test results are affected by item heterogeneity, meaning that they might measure different traits.

Total-item correlation values for the ENVST range between -.06 and .64, with an acceptable overall mean ($M = .29$, $SD = .15$) for 68 items (two items elicited a perfect score). For the ROVST, total-item correlations range between -.27 and .66, with an acceptable overall mean ($M = .32$, $SD = .20$) for 64 items (six items elicited a perfect score). Finally, for the HUVST: $min = -.03$ and $max = .56$; $M = .28$; $SD = .18$.

Based on the inter-item and total-item correlation values, it becomes amenable to theorisation that these low values are an amalgamation of different issues. Certainly, the low
inter-item correlation coefficients are due to the number of items per test and/or the lack of more participants. S1 gives some credence to the negative total-item coefficients being due to cognateness. As transpires, low-level learners can capitalise on cognate affordances, however, this will not necessarily reflect in their overall score, resulting in a discrepancy between person ability and item discrimination values. This serves as evidence for the necessity of accounting for cognates and checking a posteriori item-discrimination indices. Misfitting values are analysed further in relation to RQ2.

![Figure 5.1 ENVST mean scores by frequency](image)

**Figure 5.1 ENVST mean scores by frequency**

Regarding construct validity, the VST assesses whether students can establish a form-meaning link between the item and four definitions when a non-defining contextual sentence is also provided. Expectedly, the results should resemble the theoretical profile of a language learner (Meara, 1992).

Figure 5.1 above illustrates that students overall scored better at high-frequency levels, and as the items become less frequent, their hit rate decreases. For ease of visual representation, level one and two are collapsed into one band (E1) and consistently so for all
levels and tests. The virtue of this, apart from more comprehensive confirmatory visualisations, is that in each respective frequency band the number of data-points are increased from five to ten. This means that E1 represents the first 2,000 most frequent English words and band E7 combines the 13\textsuperscript{th} and 14\textsuperscript{th} frequency levels.

Students’ lexical profile follows roughly the same trend in their L2 as well (Figure 5.2). The means in band 5, which represents the 9\textsuperscript{th} and 10\textsuperscript{th} levels are very close to band 4 on both tests. This stipulates that either some of the items on band 4 are too difficult, or some items on band 5 are too easy for participants. These differences are further investigated in the subsequent sections.

![Figure 5.2 ROVST mean scores by frequency](image)

The downward slopes indicate that words from low-frequency ranges have a lower hit rate, indicating a decreasing knowledge on these levels, thus confirming both tests’ construct validity as acceptable and in line with Meara’s theoretical profile of a language learner (1992; see also Milton, 2009). The ensuing table provides descriptive statistics for the obtained results.
Mean scores and standard deviations are reported on the different frequency bands in Table 5.1. Of significance is the observation that the Romanian mean scores gradually decrease from high-frequency bands to lower ones, albeit, on the ENVST, bands 2 and 3, and bands 5 and 6 have similar means. It should be remembered, however, that these bands have been collapsed from 2 different frequency levels. In the case of E2, or the 3k and 4k most frequent words, the means are $M = 3.35$ and 3.15, and in the case of band E3, or the 5k and 6k most frequent words, $M = 3.46$ and 3.04 respectively. This shows that the mean scores are very close to each other and thus, differentiating between students’ lexical knowledge on these levels of the ENVST might be problematic. The reason for this is perhaps entirely due to the fact that the number of items have been reduced from 10 to 5, indicating that the probability of reaching a great variability in scores is reduced. Therefore, the different bands become less powerful in differentiating between students’ VS. This was further investigated using mixed-effects logistic regression models and maximum-likelihood ratios to establish the usefulness of each seven frequency bands in predicting students’ lexical knowledge.

Initially, a baseline or null model was created from the scores, which is the outcome variable. This was subsequently extended by a random intercept accounting for variation across participants. This resulted in a significant improvement to the model $\chi^2(1) = 16.77, p < .001$. Next, the frequency levels were added as the predictor variable (fixed effect), and again, this emerged as a significant: $\chi^2(6) = 244.87, p < .001$. Lastly, a random slope was introduced to allow for variation across participants and levels as well. This final model was also significant: $\chi^2(27) = 44.91, p < .05$. The coefficient of determination for this model is
$R^2_M = .39$ and $R^2_C = .91$, meaning the fixed effects explain 39% of variance in the model, which is exactly the same if a traditional regression model was fitted to the data, and by accounting for the random effects, the variance explained by the model significantly increases. The final model is provided in the following table.

**Table 5.2 Model of ENVST scores by level**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>$\beta$</th>
<th>SE</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.667</td>
<td>0.19</td>
<td>45.46</td>
<td>***</td>
</tr>
<tr>
<td>E2</td>
<td>-2.17</td>
<td>0.27</td>
<td>-8.00</td>
<td>***</td>
</tr>
<tr>
<td>E3</td>
<td>-2.17</td>
<td>0.27</td>
<td>-8.00</td>
<td>***</td>
</tr>
<tr>
<td>E4</td>
<td>-3.20</td>
<td>0.30</td>
<td>-10.61</td>
<td>***</td>
</tr>
<tr>
<td>E5</td>
<td>-3.54</td>
<td>0.32</td>
<td>-11.20</td>
<td>***</td>
</tr>
<tr>
<td>E6</td>
<td>-3.54</td>
<td>0.25</td>
<td>-14.13</td>
<td>***</td>
</tr>
<tr>
<td>E7</td>
<td>-4.98</td>
<td>0.29</td>
<td>-16.96</td>
<td>***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.20</td>
</tr>
<tr>
<td>E2</td>
<td>1.70</td>
</tr>
<tr>
<td>E3</td>
<td>1.74</td>
</tr>
<tr>
<td>E4</td>
<td>1.96</td>
</tr>
<tr>
<td>E5</td>
<td>2.08</td>
</tr>
<tr>
<td>E6</td>
<td>1.53</td>
</tr>
<tr>
<td>E7</td>
<td>1.98</td>
</tr>
<tr>
<td>Residual</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Sig. codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5.2 confirms a significant difference between scores from high- and low-frequency bands. Thus, it can be concluded that there is a significant difference across participants and frequency levels in the model. The same procedure was followed for the ROVST. Introducing a random intercept to the baseline model resulted in a significant improvement to the model, $\chi^2(1) = 6.88, p < .01$. The addition of levels as fixed effect further improved the model: $\chi^2(6) = 366.32, p < .001$. Finally, introducing levels as a random intercept further improved the model, providing the final model, which accounts for frequency varying across individuals: $\chi^2(27) = 44.91, p < .05$. Effect sizes are: $R^2_M = .53$ and $R^2_C = .95$, implying an explained variance of at least 53%.
Table 5.3 Model of ROVST scores by level

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>β</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.44</td>
<td>0.12</td>
<td>78.59</td>
<td>***</td>
</tr>
<tr>
<td>R2</td>
<td>-1.28</td>
<td>0.17</td>
<td>-7.51</td>
<td>***</td>
</tr>
<tr>
<td>R3</td>
<td>-1.85</td>
<td>0.26</td>
<td>-7.25</td>
<td>***</td>
</tr>
<tr>
<td>R4</td>
<td>-4.04</td>
<td>0.22</td>
<td>-18.13</td>
<td>***</td>
</tr>
<tr>
<td>R5</td>
<td>-3.61</td>
<td>0.24</td>
<td>-14.96</td>
<td>***</td>
</tr>
<tr>
<td>R6</td>
<td>-4.63</td>
<td>0.24</td>
<td>-19.07</td>
<td>***</td>
</tr>
<tr>
<td>R7</td>
<td>-4.96</td>
<td>0.25</td>
<td>-20.13</td>
<td>***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.68</td>
</tr>
<tr>
<td>R2</td>
<td>0.96</td>
</tr>
<tr>
<td>R3</td>
<td>1.69</td>
</tr>
<tr>
<td>R4</td>
<td>1.42</td>
</tr>
<tr>
<td>R5</td>
<td>1.57</td>
</tr>
<tr>
<td>R6</td>
<td>1.58</td>
</tr>
<tr>
<td>R7</td>
<td>1.62</td>
</tr>
<tr>
<td>Residual</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Sig. codes: *** p < 0.001, ** p < 0.01, * p < 0.05

Table 5.3 confirms that there is a significant difference between ROVST bands. It is of importance that since only one predictor variable (frequency levels) is involved in each model, they cannot be affected by multicollinearity. These results confirm the findings in S1 and other studies (e.g. Milton, 2007, 2009; Schmitt et al., 2001; Schmitt, 2010), and theoretically, provide some indication that the tests measure an underlying trait that closely approximates what happens in reality. Students are able to recognise quotidian lexis more accurately and struggle with low-frequency words. To determine how well the test items achieve this, the tests’ content validity was explored.

It has been articulated in S1 that the VLT and RomVLT revealed no significant difference regarding proficiency in the L2 and L3. This was despite the fact that participants are or should be considered Hungarian-Romanian bilinguals, and therefore, their L2 knowledge should be better than their L3 knowledge due to living in Romania. As this aspect adds to the validity of the tests, now I investigate this.
Table 5.4 Descriptive statistics for the overall scores

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENVST</td>
<td>54</td>
<td>41.07</td>
<td>8.46</td>
<td>19</td>
<td>59</td>
</tr>
<tr>
<td>ROVST</td>
<td>54</td>
<td>45.74</td>
<td>7.88</td>
<td>25</td>
<td>59</td>
</tr>
<tr>
<td>HUVST</td>
<td>54</td>
<td>66.24</td>
<td>2.78</td>
<td>59</td>
<td>70</td>
</tr>
</tbody>
</table>

First, the mean results in Table 5.4 indicate that participants scored better and more consistently on the ROVST in comparison to English. In order to determine whether this difference is significant, the necessary assumptions for normal probability distribution were checked for both the ENVST ($W = 0.98, p = .62$) and the ROVST ($W = .97, p = .30$), and confirmed as being tenable. Thus, the difference between the scores on the two tests was checked using a paired-samples *t*-test, which indicates that the difference is statistically significant ($t(1, 53) = -4.63, p < .001; R^2 = .54$). In line with the expectations, the difference between the L1 and the L3/L2 is obvious from the very high difference in means and the very consistent scores on the HUVST.

If the means are turned into percentages, the results confirm that 59% and 65% of items were answered correctly on the ENVST and ROVST respectively, indicating that participants on average do not reach NS levels in their L2 and L3, while in Hungarian they reached 95%. Contrary to the VLT, however, these results can be turned into VS estimations.

To translate the scores into VSs, it should be remembered that the VST employs a 1:100 sampling rate at 14 levels. Here, however, this rate had to be reduced by half, meaning that the sampling rate dropped to 1:200. In other words, the scores obtained need to be multiplied by 200. Thus, the VS estimates are in range 3,800-11,800 with $M = 8,200$ on the ENVST, and 5,000-11,800 with $M = 9,200$ on the ROVST. The lowest Romanian VS of 5,000 words and the means suggest that students’ lexical knowledge is more advanced in their L2. In comparison to S1, this tentatively indicates that due to the fact that the VST comprises 14 frequency levels, not just four bands plus the UWL, results in a more accurate test from a validity point of view. As expected, the range is considerably smaller in the L1: $min. = 11,800$ and $max. = 14,000, M = 13,200$. 

\[ \text{min.} = 11,800 \text{ and } \text{max.} = 14,000, M = 13,200. \]
Another feature of a valid vocabulary test is the ability of distinguishing between proficiency levels. In order to explore this, ANOVAs with contrasts were conducted based on year of study as possible objective indicators of proficiency. It is expected that as students progress at university, a difference would emerge between the three distinct years.

The boxplots in Figure 5.3 present the distributions of scores in the three years on the two tests. It seems that there is a difference between first and second-year students ($M = 38.87$ and $38.24$; $SD = 7.22$ and $8.64$, $n = 23$ and $17$ respectively) in comparison to third-year students ($M = 48.14$; $SD = 6.19$; $n = 14$), but there is no clear difference between the first two years of study based on the ENVST scores. Since the Levene’s test stipulates that the assumption of homogeneity of variance is tenable ($F(2, 51) = 0.93$, $p = .40$), and based on the Shapiro-Wilk test, normal distribution can be assumed ($W = 0.98$, $p = .61$), an ANOVA using contrasts will be consulted.

![Figure 5.3 Mean results on the ENVST and ROVST by year](image)

The ANOVA confirms that the scores on the ENVST significantly increase from year one and two to year three, and that year is a significant predictor of the model ($F(2, 51) =$
8.50, \( p < .001; R^2_{adj} = .22, \) medium effect). The contrasts between years suggest that the
difference between the first and third years is significant (\( t(2, 51) = 2.1, p = .04; \) Bonferroni-
adjusted \( p = 0.01 \)). The effect is also significant between the second and third years (\( t(2, 51) 
= 3.7, p < .001, \) Bonferroni-adjusted \( p = 0.01 \)). The fact that there is no difference between
the first and second-year students (\( M = 38.87, n = 23 \) and \( M = 38.24, n = 17 \)) is due to the
fact that the model only explains 22% of variance, so the other differences can derive from
individual dissimilarities, hours of exposure, baseline VS at the start of university,
motivation, etc., the effects of which might not be accentuated until students’ third year of
study.

Even though there was no expectation that the Romanian scores would be influenced
by year of study, for comparison purposes the boxplots on Figure 5.3 can be consulted. It is
apparent that the difference between the first two years is minimal (\( M = 44.83 \) and 43.59;
\( SD = 6.62 \) and 9.70; \( n = 23 \) and 17 for year one and two respectively) and the scores for year
three are slightly higher (\( M = 48.80, SD = 7.19, \) and \( n = 15 \)). In order to check whether there
is a significant effect, initially, the necessary assumptions are offered. Relying on the
assumption of homogeneity of variance (\( F(2, 51) = 1.80, p = .18 \)) and normal distribution
(\( W = 0.98, p = .35 \)), it can be inferred that these are tenable. However, the ANOVA indicates
no significant difference between years on the ROVST (\( F(2, 51) = 2.89, p = .06; R^2_{adj} = .06 \)).
Furthermore, albeit regarding the English scores, a trend analysis confirms that overall
students’ lexical knowledge increases by year (\( t(51) = 3.66, p < .001; R^2 = .46 \)), this cannot
be said for the Romanian scores (\( t(51) = 1.95, p = .06; R^2 = .22 \)). This is congruent with the
expectations as students’ VS in English is bound to increase by year three of their studies
and, apparently, the VLT is sensitive enough to capture this. Speculatively, given a more
balanced design and increased group numbers in each year, the first- and second-year groups
would possibly emerge as distinct as well. However, since not all students receive formal
instruction in Romanian anymore, their Romanian VS is not expected to increase by year of
study. Consequently, the tests perform according to the expectations as they differentiate between students’ lexical progress from their first year of study to the third in English, but not in Romanian.

Criterion validity was checked by investigating usefulness of the receptive tests in predicting academic success. First Spearman’s rank correlations were used (the GPA variable violates normality and is considered a scale) to establish whether there is a linear relationship between VS and the GPAs. These indicate that there is a statistically significant correlation between the English scores and the end-of-year average when the test was taken $\rho = .56, p < .001, N = 49$ and between the scores and the three-year GPA too: $\rho = .41, p < .05, N = 45$. The difference in the number of the samples comes from the fact that not all students reached the third year. Correlations between the Romanian scores and the two academic variables are much weaker or non-existent: $\rho = .34, p < .05, N = 49$ for the end-of-year average and $\rho = .16, p = .30, N = 45$ for the GPA upon completion.

To go beyond the correlations and gain a deeper insight into this relationship, a regression model was fitted to the data, in which GPA (only the end of year average was utilised in this analysis as between the two indicators there is a robust correlation $r = .92$, and the number of observations is higher in this one, $N = 49$) was used as the outcome variable and the ENVST and ROVST as predictor variables. Three separate models were built. In the first block, GPA was predicted by the ENVST scores: $F(1, 47) = 22.16, p < .001, R^2_{adj} = .31$, suggesting that English scores are a significant predictor of academic performance, and model explains 31% of the variance. The second block comprised the ROVST scores as predictor: $F(1, 47) = 5.23, p < .05, R^2_{adj} = .08$, revealing that the Romanian scores explain only 8% of the variance. The third model, included the ENVST and ROVST scores, and their interaction as predictors, however, neither the interaction between the predictor variables nor the fixed effects were significant and overall the model was affected by multicollinearity.
These results suggest that English VS constitutes a better predictor of academic performance than Romanian, albeit the reason for this may be that participants are studying for a degree in English. Nevertheless, it appears that comparatively both test results show positive signs of external validity. Tentatively, lexically more apt students are likely to reach higher degree classifications, which confirms previous studies (e.g. Daller and Wang, 2016; Milton and Treffers-Daller, 2013).

This section collated evidence for the reliability and validity of the test results. Some further investigations regarding validity will be offered in the ensuing sections. However, relying on the tentative evidence that the elicited scores are reliable and the tests indicate some degree of validity, I now turn to the proposed RQs.

5.8.2 The effect of explicit instruction on cognates (RQ1)

The issue of concern in this section is whether an awareness-raising exercise of cognates in the case of adult and advanced language learners facilitates cognate recognition on the two tests. To establish that a potential difference can be ascribed to the intervention and is not an emergent property of other latent factors, the homogeneity of the population is examined.

To espouse this enquiry, two objective variables can be used, which supposedly cannot be affected by the intervention per se. One is the GPA, described above, and the other one is age of acquisition of the FLs. The nature of these variables (non-normal distribution and scale variables) require non-parametric tests. The Wilcoxon rank-sum test indicates that according to the GPA variable, the Study (\(M = 7.99, SD = 2.16, N = 32\)) and the Control (\(M = 8.19, SD = 1.43, N = 28\)) groups in this study derive from the same population (\(W = 292.5, p = .30\)). Parallel to this, the age of onset of learning Romanian (\(W = 292, p = .28\); SG: \(M = 5.36, SD = 1.94, N = 22\); CG: \(M = 4.75, SD = 2.16, N = 32\)) and English (\(W = 399, p = .41\); SG: \(M = 7.68, SD = 2.73, N = 22\); CG: \(M = 8.59, SD = 3.50, N = 32\)) both emerge non-distinct across the two groups. This stipulates that the participants in the experimental groups
have been indeed randomly assigned to either the Study or Control groups (SG/CG).

### Table 5.5 Descriptive statistics for the ENVST based on groups and levels

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Total</th>
<th>Cog</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>E5</th>
<th>E6</th>
<th>E7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>M</td>
<td>42.18</td>
<td>24.73</td>
<td>8.64</td>
<td>6.36</td>
<td>6.91</td>
<td>5.73</td>
<td>5.50</td>
<td>5.27</td>
<td>3.77</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>9.65</td>
<td>4.54</td>
<td>1.76</td>
<td>2.17</td>
<td>1.41</td>
<td>2.03</td>
<td>2.20</td>
<td>1.39</td>
<td>2.00</td>
</tr>
<tr>
<td>Control</td>
<td>M</td>
<td>40.31</td>
<td>23.31</td>
<td>8.69</td>
<td>6.59</td>
<td>6.22</td>
<td>5.28</td>
<td>4.88</td>
<td>5.03</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>7.60</td>
<td>3.86</td>
<td>1.12</td>
<td>1.76</td>
<td>2.07</td>
<td>2.23</td>
<td>2.06</td>
<td>1.40</td>
<td>1.74</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>41.07</td>
<td>23.72</td>
<td>8.67</td>
<td>6.50</td>
<td>6.50</td>
<td>5.46</td>
<td>5.13</td>
<td>5.13</td>
<td>3.69</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>8.46</td>
<td>4.19</td>
<td>1.40</td>
<td>1.92</td>
<td>1.85</td>
<td>2.14</td>
<td>2.12</td>
<td>1.39</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Table 5.5 reports descriptive statistics for the ENVST, based on the experimental conditions. It emerges that the SG reached a higher mean score than the CG. The Shapiro-Wilk test for normality indicates that the assumption for normal distributions is tenable in the case of the SG ($W = 0.97$, $p = .62$) and the CG ($W = 0.96$, $p = .46$) as well. Levene’s test confirms that the assumption of homogeneity of variance is also tenable between the two groups: $F(1, 52) = 1.99$, $p = .16$. Consequently, to explore whether the difference between the means is significant, a two-sample *t*-test was conducted, which indicates no difference between the groups: $t(38.04) = -0.76$, $p > .05$; $R^2 = .12$.

### Table 5.6 Descriptive statistics for the ROVST based on groups and levels

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Total</th>
<th>Cog</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>M</td>
<td>45.32</td>
<td>26.41</td>
<td>9.41</td>
<td>8.05</td>
<td>7.18</td>
<td>5.14</td>
<td>5.86</td>
<td>5.18</td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>9.28</td>
<td>4.96</td>
<td>0.85</td>
<td>1.53</td>
<td>2.38</td>
<td>1.98</td>
<td>1.64</td>
<td>1.56</td>
<td>1.82</td>
</tr>
<tr>
<td>Control</td>
<td>M</td>
<td>46.03</td>
<td>26.19</td>
<td>9.47</td>
<td>8.25</td>
<td>7.88</td>
<td>5.59</td>
<td>5.81</td>
<td>4.56</td>
<td>4.47</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>6.91</td>
<td>3.88</td>
<td>0.92</td>
<td>1.1</td>
<td>1.74</td>
<td>1.52</td>
<td>1.73</td>
<td>2.05</td>
<td>1.85</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>45.74</td>
<td>26.28</td>
<td>9.44</td>
<td>8.17</td>
<td>7.59</td>
<td>5.41</td>
<td>5.83</td>
<td>4.81</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>7.88</td>
<td>4.32</td>
<td>0.88</td>
<td>1.28</td>
<td>2.03</td>
<td>1.72</td>
<td>1.68</td>
<td>1.87</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Table 5.6 reveals that the difference between the groups overall is even smaller on the ROVST. Merely out of consistency reasons, the difference will be checked using a two-sample *t*-test. The normality test indicates that the assumption has been met (SG: $W = 0.95$, $p = .25$; CG: $W = 0.98$, $p = .84$), just as the assumption for equal variances ($F(1, 52) = 2.14$, $p = .15$). As expected, the difference between the two means is not significant: $t(36.45) = 0.31$, $p > .05$ with a small effect size ($R^2 = .05$).
In relation to the cognate items, the scores on both tests indicate that the means are very close to each other and the t-tests indicate no differences.

In conclusion to RQ1, the data analysis substantiates that the EI and the awareness-raising task on cognates produced no effect either on the total scores or on the cognates. Tentatively, this can also suggest that either the lexical test is not sensitive enough to be able to detect the effect of such short awareness-raising exercises or cognate recognition in adults and proficient learners is automatic.

**5.8.3 Exploring the relationship between the English and Romanian lexica (RQ2)**

RQ2 is addressed in three parts. From a within-subject perspective, part A investigates the relationship between overall lexical knowledge and cognate scores in the L2 and L3. Following this, part B explores the facilitatory effect of cognates in a by-item analysis. Part C, then, examines the interplay between cognateness and word frequency, and whether this is influenced by lexical proficiency.

### A. Comparing vocabulary size in English and Romanian

![Figure 5.4 Scatterplot with overall scores on the two tests](image-url)
The arguments in S1 posited a strong relationship between the languages under study. To verify this in S2, first, I explored the existence of a linear relationship between the overall scores. This is presented in Figure 5.4, which confirms that the higher the VS in one language, the higher it is likely to be in the other with some individual differences. Pearson’s correlation emerges as statistically significant between the VSs: $r = .59$ ($p < .001$, $N = 54$ in all cases, unless otherwise stated).

Table 5.7 reports Spearman’s rank correlation values for the individual levels as some of these variables have violated the assumption of normality. It indicates mostly weak correlations, except for bands E4 and R4. For instance, between the first 2k most frequent English (E1) and Romanian (R1) words tested, the correlation is $\rho = .27$, $p = .05$, which is a weak correlation. The reason for this difference between S1 and S2 potentially lies in the fact that while one band on the VLT tested 18 items, here only 10 items are tested within a (collapsed) band. This provides a perhaps less accurate sample of items from the respective frequency bands and this might also explain the lower correlation for the total scores.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>$\rho$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1 ~ R1</td>
<td>.27</td>
<td>.05 *</td>
</tr>
<tr>
<td>E2 ~ R2</td>
<td>.10</td>
<td>.49</td>
</tr>
<tr>
<td>E3 ~ R3</td>
<td>.31</td>
<td>.02 *</td>
</tr>
<tr>
<td>E4 ~ R4</td>
<td>.49</td>
<td>***</td>
</tr>
<tr>
<td>E5 ~ R5</td>
<td>.14</td>
<td>.31</td>
</tr>
<tr>
<td>E6 ~ R6</td>
<td>.26</td>
<td>.05 *</td>
</tr>
<tr>
<td>E7 ~ R7</td>
<td>.32</td>
<td>.02 *</td>
</tr>
</tbody>
</table>

Sig. codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

If the scores are separated into cognate and non-cognate scores, the results indicate a moderate correlation of $r = .63$ (Pearson’s correlation) for cognates and a weak correlation of $r = .31$, $p = .02$ for non-cognate scores between the two languages. This shows that cognates show a more robust linear relationship between the languages than non-cognate scores. Additionally, both cognate and non-cognate scores in English seem to correlate
highly with the overall English scores: $r = .90$ and .93 respectively. Similarly, in Romanian, both cognate and non-cognate scores correlate equally well with the total scores: $r = .89$ and .90. The results also confirm that the higher the cognate score is in one language, the larger the VS is in the other: between L2 cognates and L3 total score $r = .64$, and between L3 cognates and L2 total score $r = .62$.

To complement the correlation analysis carried out hitherto, linear regression models were employed. In order to establish whether L2 VS is a significant predictor of L3, ENVST total scores were used as an outcome variable and, in the first instance, the ROVST scores as predictors. This led to a significant model: $F(1, 52) = 28, p < .001, R^2_{adj} = .34$, which means that the L2 scores account for 34% of variance, a medium effect size. Next, the interest fell on whether either the L2 cognate or non-cognate scores, or the interaction between the two, prove to be a significant predictor of the model. Since neither the interaction nor the main fixed effects in conjunction seemed to significantly predict the English scores, in order to avoid multicollinearity, two separate models have been built. Both the Romanian cognate ($F(1, 52) = 35.87, p < .001, R^2_{adj} = .41$) and non-cognate scores ($F(1, 52) = 11.29, p < .01, R^2_{adj} = .16$) have a significant effect, however, cognates apparently explain more variance in the L3 overall scores. Compared to the overall Romanian scores, which explain 34% of the variation in the English scores, Romanian cognate scores seem to be a better predictor with 41% of variance explained.

The analysis shows that fitting a model to the overall ROVST scores with the ENVST cognate and non-cognate scores as predictors also yields significant models: $F(1, 52) = 32.8, p < .001, R^2_{adj} = .38$ and $F(1, 52) = 15.54, p < .001, R^2_{adj} = .22$. Congruent with the findings in S1 and the above, cognates emerge as better predictors of VSs than non-cognates. These significant findings will be further discussed in Chapter 6.
B. The effect of cognate facilitation on testing

This section investigates the relationship between the languages under study further. Particularly, it explores the effect of cognates on lexical tests. S1 revealed that in English more cognates were recognised than non-cognates, suggesting a cognate facilitation effect. However, this was not the case in Romanian as cognates and non-cognates had roughly equal means. In order to determine the existence of a cognate facilitation effect in the present study, first, the descriptive statistics are reported.

Table 5.8 summarises the percentage of correct answers to cognates and non-cognates on the three tests. The mean item facility indices for the English and Romanian tests indicate that the 39 cognate items were answered correctly by more students (61 and 67%) than the 31 non-cognates (56 and 63%). However, in their L1, participants’ responses suggest that slightly more non-cognate items were answered and more consistently so than cognate items.

<table>
<thead>
<tr>
<th>Test</th>
<th>Cognate Status</th>
<th>N</th>
<th>Mean Item Facility in %</th>
<th>Item facility range</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>ENVST</td>
<td>Cognate</td>
<td>39</td>
<td>60.82</td>
<td>17</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Non-cognate</td>
<td>31</td>
<td>55.94</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>70</td>
<td>58.66</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>ROVST</td>
<td>Cognate</td>
<td>39</td>
<td>67.31</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Non-cognate</td>
<td>31</td>
<td>62.84</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>70</td>
<td>65.33</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>HUVST</td>
<td>Cognate</td>
<td>39</td>
<td>93.79</td>
<td>43</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Non-cognate</td>
<td>31</td>
<td>95.61</td>
<td>76</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>70</td>
<td>94.60</td>
<td>43</td>
<td>100</td>
</tr>
</tbody>
</table>

To explore whether the difference between the means is substantial for cognate and non-cognate items, the necessary assumptions were tested for in the first instance. Since the assumption for normal distributions have been violated with regard to some variables, the non-parametric Wilcoxon rank-sum test was applied. This indicates a non-significant difference on the English test ($W = 687.5, p = .32, R^2 = .12$), and the Romanian test ($W = 675, p = .41, R^2 = .10$) as well with small effect sizes. The difference is also non-significant in the case of the Hungarian test ($W = 611.5, p = .93, R^2 = .01$).
Figure 5.5 ENVST Item facility boxplot for the different levels by cognate status

Figure 5.6 ROVST Item facility boxplot for the different levels by cognate status

Figures 5.5 and 5.6 facilitate the comparison of scores to cognate and non-cognate
Results

items on each frequency band. With regard to the ENVST, it can be observed that the most frequent words (E1), due to an approximate ceiling effect, show no visible difference between cognate and non-cognate facility indices. In the case of E2, E4 and E5, however, there is a clear facilitative effect of cognates, whereas on the lower bands the advantage of cognates seems to diminish. Band E3 apparently stands out as non-cognates were answered correctly by more students than cognates. The reason for this is that there are only three non-cognate items on this band and they seem to elicit more responses than the seven cognate items on the same band. On E7 the words are from such low-frequency levels that cognateness apparently does not have an effect. This serves as evidence that even though cognate words might not bear a facilitative effect that is significant for the overall scores, depending on frequency, cognates might be more easily recognised than non-cognates in English.

Similarly to the ENVST, the responses to the first 2k Romanian words (R1) indicate that there is no difference between the means, however, cognate scores might be more consistently answered (Figure 5.6). Bands (R) 2, 3, 5, and 6 depict a clear facilitative effect of cognates. Band 7 stands out on both tests as there seems to be no clear difference between cognate and non-cognate scores, which is a tentative indication that in the case of very low-frequency words, frequency plays a more crucial role than cognate status.

In sum, although the total mean item facility indices suggest that there is no clear facilitative effect of cognates, based on the individual frequency levels, there is some evidence that cognates elicit a higher hit rate in the case of mid-frequency ranges (from 3k to 12k in this case, based on the results), and on the last band (13-14k) items are perhaps so rare that cognate status becomes a minor factor. These differences will be further investigated in the following sections.

In the validity section, the results suggested that some of the bands might not be able to discriminate well between proficiency. Furthermore, Table 5.8 demonstrates that the
maximum percentage of correct answers was 100, suggesting that one or more items elicited a ceiling effect. The following succinct item analysis will help to explore whether this is due to cognacy or misfitting items in terms of ambiguous definitions, or other discrepancies. It is worth mentioning that the English ability estimates indicate that no participant reached a perfect score and the estimates ranged from .27 to .84 ($M = .59, SD = .12$). In Romanian, the lowest ability score was .36 and the highest was .84 ($M = .65, SD = .11$). The item facility indices, as confirmed by Table 5.8 as well, in English ranged from .09 to 1 ($M = .59, SD = .25$) and on the ROVST from .07 to 1 ($M = .65, SD = .28$).

Although in assessments, generally the $>.30$ and $<.70$ range is applied for selecting appropriate items, due to the fact that in both studies items are selected according to their frequencies, and participants are adult students majoring in English and living in Romania, a greater upper and lower limit for item facility is accepted. Thus, only items that barely generate any variance, or in other words cannot differentiate between students reliably, are further investigated.

![Figure 5.7 Item Facility distribution by language and cognate status](image)

Figure 5.7 Item Facility distribution by language and cognate status
By inspecting Figure 5.7, first, it can be noticed that the majority of items (42 English and 45 Romanian out of 70) fall above the .50 facility mark, which serves as evidence that the items overall can be considered easy. If this is broken down into cognate and non-cognate items, it turns out that on the ENVST 26 cognates and only 16 non-cognates fall above the .50 mark, whereas on the ROVST this figure is much closer with 24 cognates and 21 non-cognates. Moreover, the upper quartile of the facility index suggests that 13 cognates versus eight non-cognates from the English test, and 20 cognates versus 11 non-cognates on the Romanian test are above .75. This means that overall cognates elicited more answers and can be thus regarded as easier items.

Next, items that elicited a perfect or near perfect score are inspected. The selection criterion for this equals that they had an item difficulty index of > .9 (i.e. more than 90% of students provided a correct answer to that item). Therefore, these can be considered as misfitting items. Only two ENVST items generated a perfect score: shoe, which is a non-cognate from the first 2k most frequent words (E1) and dinosaur (E2), which is a multilateral cognate between all three languages.

Eight other items had a facility index > .90: see, poor, period, and microphone (E1); soldier (E2); palette (E4); octopus (E5); and puma (E6). The item facility boxplots have already indicated that on the first band, there is no cognate effect, therefore, these items have a high facility index simply because of their high frequency. On the other bands, however, except the item soldier, which is only cognate in Romanian and English, all items are multilateral cognates. This gives the impression that due to cognateness, for Romanian learners, these lexemes become misfitting items generating too high response ratios and should be reconsidered in further tests. Another possible reason, could be that some of these words are semantically related to animals, which have the tendency of being acquired quite early by learners.

Despite the fact that no English items induced a flooring effect, the analysis now turns
to items that have a facility index below .25, as these are considered to be too difficult and therefore, cannot discriminate well between students’ abilities. There are seven items that fit this criterion: veer (E3); quilt (E4); cranny (E5); hessian and refectory (E6); ubiquitous and atoll (E7). No misleading or ambiguous definitions have been identified for these items. Some incorrect definitions attracted more responses though that could not be defined as a random choice. One of these is the incorrect answer for quilt: *feather pen, which is the actual definition of quill, and the other one is refectory: *room with glass walls for growing plants, which is the definition of conservatory. Although these items share some orthographic or thematic similarities, the actual definitions cannot be surmised ambiguous or misleading. Consequently, it can be speculated that by design, the items and definitions test recognition of the actual word. Moreover, regarding the cognate status of these items, except quilt and cranny, they are either bi- or multilateral cognates and thus the very low hit rate can be regarded as further proof that these were simply not part of students’ lexicon.

In the case of the ROVST, there are six items that generated a perfect score: cer (sky), cal (horse), rece (cold), and memorie (memory) (R1); farmacie (pharmacy) and șofer (chauffeur) (R2). As has been inferred previously, cognateness does not have an effect on band one, so it can be assumed that the three non-cognates and one cognate item were simply too easy for advanced students, as one might expect. The items from R2, however, are both cognates, which could have played a (partial) role in the fact that they facilitated a perfect score. Disregarding the first frequency band, which featured three other items > .90, there were nine items in total that elicited correct responses from over 90% of testees: ironie (irony), biografie (biography), adult (adult) (R2); bronz (bronze), răni (hurt), glucoză (glucose) (R3); păgân (pagan) (R4); variație (variation) (R6); disertație (dissertation) (R7). It becomes evident from these words that the high hit rate these words have facilitated is majorly due to either being bi- or multilateral cognates. Răni is the exception to this as a non-cognate but quite commonly used in everyday settings.
Further five words have elicited answers from less than 25% of participants. Two of these items are non-cognates: *cochilie* (*shell, R6*) and *destoinic* (*capable, R7*). No distractors facilitated an unusually high number of responses for these items and due to their low frequency, it can be argued that these words were simply unknown to students. Three items were cognates: *monopol* (*monopoly, R4*), *reitera* (*reiterate, R6*) and *nihilism* (*nihilism, R7*). Surprisingly, the first item, which is a multilateral cognate, elicited only six correct answers, meaning that it has an item facility of .11. This, taken together with the fact that it is an item from the fourth band, suggests that the item might be problematic. A close inspection of the distractors and answers indicates that most students chose as definition the distractor *guvernat de un singur partid* (*governed by one single party*) and the reason for this could be that the non-defining example sentence was: *Statul are monopol* (*The state has monopoly*). Although it can only be speculated, the word *state* might have served as a strong prime for *govern* in the definition and, therefore, the item became misleading.

The other lexeme, *reitera*, elicited even fewer answers with an item facility index of .07, despite the fact that it is a cognate in English (*reiterate*). Again, the anomaly is a combination of distractor and example sentence interaction. The non-contextual sentence can be translated as *The text reiterated the idea* (*Textul reitera ideea*) and the distractor that generated most answers was *clarifica* (*clarify*), instead of *avea loc încă o data* (*take place again*), which partially explains why students were misled by the ambiguous distractor *clarifica*. With regards nihilism, the fact that it is a cognate, but still generated a low response rate lies in the fact that it is a very low-frequency item and students, thus, might be less aware of its exact meaning. The distractor *credinţa religioasă după Protestantism* (*religious belief after Protestantism*) cannot be considered confusable with *negarea radicală a unui sistem de valori* (*radical negation of a system of values*), which is the correct definition.

To summarise the brief item evaluation above, it can be asserted that except for the highest and lowest frequency bands, cognateness becomes an issue for vocabulary testing.
Since cognates are overlapping in meaning and form, there is a higher probability of answering these correctly, which results in low item discrimination values. In the cases in which cognate items had a very low facility index, it actually turned out that the items were either misleading or had an ambiguous distractor due to the interaction with the example sentence. This is of significance as it brings to light further evidence why the cognate status of words should be an important factor when constructing vocabulary tests, and it also confirms S1.

C. Lexical proficiency and cognate recognition

S1 has indicated that cognate recognition is proficiency dependant with a large magnitude. That is, lexically more apt students recognise significantly more cognates than their peers with a lower VSs, and this will stand in both the L2 and L3. As before, the median split is used on the overall scores to divide participants by lexical proficiency.

Table 5.9 ENVST cognate results on different frequency levels

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Total</th>
<th>Cog total</th>
<th>CE1</th>
<th>CE2</th>
<th>CE3</th>
<th>CE4</th>
<th>CE5</th>
<th>CE6</th>
<th>CE7</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>M</td>
<td>47.68</td>
<td>26.68</td>
<td>3.71</td>
<td>4.46</td>
<td>4.68</td>
<td>3.86</td>
<td>2.50</td>
<td>4.50</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.99</td>
<td>2.72</td>
<td>0.53</td>
<td>0.64</td>
<td>1.28</td>
<td>0.76</td>
<td>0.58</td>
<td>1.07</td>
<td>1.23</td>
</tr>
<tr>
<td>Low</td>
<td>M</td>
<td>33.96</td>
<td>20.54</td>
<td>3.50</td>
<td>3.38</td>
<td>3.31</td>
<td>2.77</td>
<td>1.69</td>
<td>3.81</td>
<td>2.08</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.83</td>
<td>2.98</td>
<td>0.76</td>
<td>0.80</td>
<td>1.32</td>
<td>1.31</td>
<td>0.79</td>
<td>1.30</td>
<td>1.38</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>41.07</td>
<td>23.72</td>
<td>3.61</td>
<td>3.94</td>
<td>4.02</td>
<td>3.33</td>
<td>2.11</td>
<td>4.17</td>
<td>2.54</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>8.46</td>
<td>4.19</td>
<td>0.66</td>
<td>0.90</td>
<td>1.46</td>
<td>1.18</td>
<td>0.79</td>
<td>1.22</td>
<td>1.37</td>
</tr>
</tbody>
</table>

Table 5.9 reports means and standard deviations for the total scores, total cognate scores, and cognate scores on the different bands on the ENVST separated into High and Low performing students. Firstly, it can be noticed that the test can distinguish between the proficiency groups quite well. The High group ($M = 47.68$, $SD = 4.99$) outperformed the Low group ($M = 33.96$, $SD = 4.83$) but there is no difference in consistency based on the standard deviations. Due to the fact that the Low group’s scores violate the assumption of normal distributions, the difference in the total scores is investigated using the Wilcoxon rank-sum test. This reveals that the mean total score of the High group is significantly larger.
than the Low group’s mean: $W = 728, p < .001$, with a large magnitude $R^2 = .86$.

Following this, the means suggest that based on the entirety of the cognate scores, the group with the larger VS ($M = 26.68, SD = 2.72$) recognised more cognates than the other group ($M = 20.54, SD = 2.98$). This is also confirmed by the Wilcoxon test: $W = 690, p < .001, R^2 = .77$.

Figure 5.8 ENVST mean cognate score interaction by proficiency

Figure 5.8 reports the interaction between the two groups based on word frequency. As has been mentioned earlier, the first band (E1) is not entirely reliable in distinguishing between cognate and non-cognate knowledge, or between high and low performers based on their cognate knowledge (CE1). However, the graph illustrates that the difference between the High and Low groups preponderates on bands CE2 to CE4 inclusively, and then, it moderately stabilises for lower frequency items. This suggests that the larger the VS overall is, the greater participants cognate knowledge is on the mid-frequency levels, albeit this advantage diminishes at lower frequency levels.
Table 5.10 ROVST cognate results on different frequency levels

<table>
<thead>
<tr>
<th>Test</th>
<th>N</th>
<th>Total</th>
<th>Cog total</th>
<th>CR1</th>
<th>CR2</th>
<th>CR3</th>
<th>CR4</th>
<th>CR5</th>
<th>CR6</th>
<th>CR7</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>28</td>
<td>51.82</td>
<td>29.14</td>
<td>3.86</td>
<td>5.00</td>
<td>6.11</td>
<td>3.14</td>
<td>1.89</td>
<td>5.25</td>
<td>3.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.17</td>
<td>2.95</td>
<td>0.36</td>
<td>0.00</td>
<td>0.69</td>
<td>0.97</td>
<td>0.69</td>
<td>1.40</td>
<td>1.37</td>
</tr>
<tr>
<td>Low</td>
<td>26</td>
<td>39.19</td>
<td>23.19</td>
<td>3.81</td>
<td>4.77</td>
<td>4.65</td>
<td>2.38</td>
<td>1.77</td>
<td>3.50</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.19</td>
<td>3.31</td>
<td>0.49</td>
<td>0.51</td>
<td>1.65</td>
<td>1.06</td>
<td>0.59</td>
<td>1.21</td>
<td>1.01</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>45.74</td>
<td>26.28</td>
<td>3.83</td>
<td>4.89</td>
<td>5.41</td>
<td>2.78</td>
<td>1.83</td>
<td>4.41</td>
<td>3.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.88</td>
<td>4.32</td>
<td>0.42</td>
<td>0.37</td>
<td>1.43</td>
<td>1.08</td>
<td>0.64</td>
<td>1.57</td>
<td>1.44</td>
</tr>
</tbody>
</table>

The same descriptive statistics for the ROVST are reported in Table 5.10. The overlap between participants in the High and Low groups in the two languages is 70%, which means that 38 out of 54 students were assigned to exactly the same group in the two languages. Overall the test is capable of distinguishing between students with differing Romanian proficiency. This is confirmed by the fact that the High group ($M = 51.82$, $SD = 4.17$) has not only outperformed the Low group ($M = 39.19$, $SD = 5.19$) but performed slightly more consistently as well. The difference is confirmed as significant by the Wilcox test: $W = 728$, $p < .001$, $R^2 = .86$. The difference between the cognate means indicates that participants with a larger Romanian vocabulary recognise significantly more cognates than their lexically less apt peers: $W = 669.5$, $p < .001$, $R^2 = .72$.

Figure 5.9 indicates that on CR1, CR2, and CR5 the difference between the two proficiency groups is minuscule. The difference on CR3 and the two low-frequency bands (CR6 and CR7), however, suggest that testees with an increased lexical dexterity recognise more cognates than the lower proficiency group. Compared to Figure 5.8 these results are more convoluted to interpret though. The reason for this could be that as the item analysis has shown, there were a couple of misfitting Romanian items, resulting in learners’ erratic responses in both groups regardless of cognates. Moreover, it also pertains that cognates are not equally distributed on levels.
Figure 5.9 ROVST mean cognate score interaction by proficiency

Although there is no statistically significant difference between the High and Low groups’ mean item facility indices based on the cognate status of items, it is interesting to notice that the Low group in English recognised more cognates ($M = 52.66$, $SD = 27.75$) than non-cognates ($M = 43.30$, $SD = 24.59$), while the High group received almost identical mean scores on cognates ($M = 68.42$, $SD = 25.29$) and non-cognates ($M = 67.74$, $SD = 24.32$) in both languages. Similarly in Romanian, the Low group recognised more cognates ($M = 59.47$, $SD = 33.17$) than non-cognates ($M = 51.61$, $SD = 27.86$), while the High group responded to the same amount of cognate ($M = 74.73$, $SD = 28.16$) and non-cognate ($M = 73.16$, $SD = 27.03$) items. The following graph illustrates just this.
With the purpose of further substantiating this finding, a mixed effect model was fitted to the data. This is a by-item analysis attempting to explore the relationship between proficiency and cognateness. For the ENVST scores, the baseline model includes the percentage of students that answered the items correctly. Introducing a random intercept for items and a random slope for proficiency results in a significant improvement in the model according to the log-likelihood ratio: $\chi^2(3) = 32.34, p < .001$. This confirms again that the scores vary across items and groups. Next, a fixed effect for proficiency is introduced, which again improves the model: $\chi^2(1) = 71.70, p < .001$. In the third step, the cognate status of items and the interaction between proficiency and cognateness have been added. This results in a significant change in the model, $\chi^2(2) = 6.65, p < .05$. As a final step, the cognate status of items was also allowed to vary across items and groups as a random slope, however, this did not yield a better model. The final model is described in the following Table (5.11).
Table 5.11 Model of ENVST scores by proficiency and cognateness

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>( \beta )</th>
<th>( SE )</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>68.41</td>
<td>3.98</td>
<td>17.18</td>
<td>***</td>
</tr>
<tr>
<td>EN.Low</td>
<td>-15.74</td>
<td>2.28</td>
<td>-6.91</td>
<td>***</td>
</tr>
<tr>
<td>NC</td>
<td>-0.66</td>
<td>5.98</td>
<td>-0.11</td>
<td>.91</td>
</tr>
<tr>
<td>EN.Low:NC</td>
<td>-8.70</td>
<td>3.42</td>
<td>-2.54</td>
<td>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>( SD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>23.43</td>
</tr>
<tr>
<td>Prof.</td>
<td>Item</td>
</tr>
<tr>
<td>Residual</td>
<td>7.17</td>
</tr>
</tbody>
</table>

Sig. codes: *** \( p < 0.001 \), ** \( p < 0.01 \), * \( p < 0.05 \)

The r-squared values for the final model indicate that the model explains at least 14% of the variance, \( R^2_M = .14 \) and \( R^2_C = .93 \). According to the VIF and tolerance values, there are no issues of multicollinearity within the model. Table 5.11 describes that the main effect of proficiency, if all other elements are ignored, is a significant predictor of the model: \( b = -15.74, t(68) = -6.19, \ p < .001 \). This confirms that compared to the High group’s scores, the Low group’s scores are significantly different. The negative slope \( (b) \) explains that the direction is negative, i.e. the Low group attains lower scores. The fixed effect of cognateness is non-significant \( b = -0.66, t(68) = -0.11, \ p > .05 \), indicating that there is no difference between cognate and non-cognate items overall. However, the higher order interaction term is significant \( (b = -8.70, t(68) = -2.54, \ p < .05 \), suggesting that compared to the High group’s cognate and non-cognate scores, the Low group attained significantly fewer hits to non-cognate items than cognates. This was visualised in Figure 5.10 above.

Introducing a random slope and intercept for ROVST items and proficiency significantly improved the baseline model: \( X^2(3) = 45.44, \ p < .001 \). Scores divided by proficiency were added as the first fixed effect, which again yielded a better model: \( X^2(1) = 53.85, \ p < .001 \). As the last step, cognateness has been introduced and the interaction between proficiency and cognateness of items. This resulted in a non-significant improvement to the model, however, to be able to compare the effects on the two tests, I retained this model:
Chapter 5: Study Two (S2)

$X^2(2) = 2.62, p > .05$. The model does not violate multicollinearity according to the VIF and tolerance coefficients and according to the effect sizes, the fixed effects in the model explain 10% variation: $R^2_M = .10$ and $R^2_C = .93$.

By contrast to the English test, Table 5.12 below reveals that neither the main effect of cognateness of items nor the interaction between proficiency and cognateness reach statistical significance. The only significant fixed effect in the model is the difference in the scores of the High and Low group: $b = -15.25, t(68) = -5.70, p < .001$. Indeed, by inspecting the interaction Figure (5.10), it can be observed that the difference in the case of the Low group between cognates and non-cognates is not as accentuated as in English. Although assuming a tendency might be appropriate.

Table 5.12 Model of ROVST scores by proficiency and cognateness

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>β</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>74.73</td>
<td>4.43</td>
<td>16.87</td>
<td>***</td>
</tr>
<tr>
<td>RO.Low</td>
<td>-15.25</td>
<td>2.68</td>
<td>-5.70</td>
<td>***</td>
</tr>
<tr>
<td>NC</td>
<td>-1.57</td>
<td>6.66</td>
<td>-0.24</td>
<td>.81</td>
</tr>
<tr>
<td>RO.Low:NC</td>
<td>-6.29</td>
<td>4.02</td>
<td>-1.56</td>
<td>.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>23.43</td>
</tr>
<tr>
<td>Prof.</td>
<td>Item</td>
</tr>
<tr>
<td>Residual</td>
<td>8.20</td>
</tr>
</tbody>
</table>

Sig. codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

One last issue to address within this section is the percentage of hits to items on the different frequency levels, based on proficiency and cognateness. S1 provided evidence that on the English test (VLT), and to some extent on the RomVLT, students in the Low-proficiency group could capitalise on the affordances of cognates. The ensuing figures allow investigating this aspect in the present case.
Figure 5.11 Interaction between proficiency and cognateness based on frequency on the ENVST.

Figure 5.12 Interaction between proficiency and cognateness based on frequency on the ROVST.
Figure 5.11 reports the distribution of hits to English cognate and non-cognate items by the two proficiency groups under study based on the frequency levels. Whilst on the most frequent items (E1), there is no noticeable difference pertaining to cognateness, on band E2 it seems that cognates elicited a higher percent of hits compared to non-cognates in the case of the Low group. In other words, by capitalising on cognate affordances, the Low group managed to reduce the difference between them and the High group, albeit this is not the case when it comes to non-cognate items. On bands E4 to E6 inclusively, it can be observed that the distance between non-cognate and cognate items is larger in the case of the Low group, in the sense that proportionally they recognised more cognates than non-cognates by contrast to the High group. This is best illustrated by E6 as there is no noticeable difference between the numbers of cognates recognised by the two proficiency groups. This lends additional substance to the findings revealed by the mixed model, namely that cognates incur as potential affordances for testees of lower proficiency. On the lowest frequency band (E7), cognateness does not seem to influence the scores for either of the groups.

The findings in S1, the ROVST (Figure 5.12 above) and ENVST scores together indicate that items from high-frequency bands (E1 and R1) and the lowest frequency bands (E7 and R7) are not influenced by the cognate status of words in either proficiency groups. This suggests that frequency potentially plays a more important role on these ‘extreme’ levels. On R2, however, it seems that the Low group are on par with the High group in terms of cognate recognition, despite the large gap on non-cognate items. On the subsequent bands (R3, R4, and R5) the difference between non-cognates and cognates is certainly not large for the Low group, however, it implies a tendency towards closing the gap on cognate items.

In summary to RQ2, the results revealed a robust relationship between lexical knowledge in the L2 and L3 with regards to the tests employed and the form-meaning mapping skill tested. Additionally, it has been reported that cognate scores are a strong predictor of overall lexical proficiency not just in the same, but the other tested language as
well. Based on item facility, it has been corroborated that although there is a tendency for answering more cognates than non-cognates on both tests, the difference between the scores does not reach significance.

The results on the seven different bands, however, suggested that there is some evidence for the facilitative effect of cognates with the exception of the highest and lowest frequency bands. The item analysis also confirms this: participants could benefit from the affordances that cognate lexemes provided, as more cognate items fall above the .50 and .75 item facility marks than non-cognates. Additionally, due to the strong lexical relationship between English and Romanian, congruent items can potentially become misfitting items and, therefore, problematic for vocabulary tests.

Finally, the results have also suggested that cognate recognition is also proficiency dependent, the larger the VS, the more cognates are recognised. However, the cognate facilitation effect is more evident in students with lower VSs. Taken together with S1, it is entirely amenable to theorisation that cognates have an impact on lexical tests and can serve as a catalyst for reducing the difference between high and low performing students.

5.8.4 Exploring speed of lexical access and vocabulary size (RQ3)

The focus is now turned to the data obtained through measuring participants’ RT in their three dominant languages. Although the website measures four values (time of the first click, time of the last click, submission time, and the number of clicks), in the ensuing sections only the submission time to correct answers are analysed. I considered this to provide the most detailed information on participants’ performance in terms of a temporal measure and to allow a direct comparison with previous studies (e.g. Tanabe, 2016).

The data analyses for RQ3 are separated into five parts. Part A starts with a multilingual perspective and compares the RTs in the three languages. Parts B and C address the size and frequency effects in relation to RT. Part D zooms in on the effect of cognates on RTs. Finally, part E addresses word difficulty beyond frequency and cognateness and
explores whether word length and POS emerge as significant descriptors of either VS or speed of lexical access.

A. Comparing RTs to vocabulary size in the three languages tested

The anticipation here is that the relative RTs of form-meaning mapping, or lexical-semantic access, are in line with lexical proficiency. That is, RTs should be the fastest in the L1, moreover, as the test scores in the above sections indicate, RT should also be faster in L2 than L3.

Based on this premise, the mean RTs to hits show that on average it took six seconds \( (\text{min.} = 0.81, \text{max} = 30.02, M = 6.39, SD = 3.44) \) to submit a response to a Hungarian item, while 13 and 10 seconds to English \( (\text{min.} = 0.07, \text{max} = 30.02, M = 12.52, SD = 6.96) \) and Romanian \( (\text{min.} = 0.36, \text{max} = 30.02, M = 9.83, SD = 5.57) \) items respectively. A paired-samples \( t \)-test between the Romanian and English RTs confirms that the difference is significant: \( t(53) = 7.83, p < .001, R^2 = .73 \), a large effect size. This difference is illustrated in Figure 5.13 below and if this is compared to the mean results in Table 5.3, it can be extrapolated that testees reached the highest score in Hungarian and they were also the fastest on average in this language. This is followed by Romanian, in which they reached the second highest scores and the mean RT also aligns with this. The lowest was the English VS and RT, meaning they were the slowest in this language. Furthermore, the dispersion values explain that RTs were most consistent in Hungarian, followed by Romanian, and then English. Consequently, the more proficient students are in a language, the faster and more accurate their responses appear to be on the test. This lends some evidence for asserting that speed of response even on a multiple-choice test format, which by design incurs less accurate estimates compared to yes/no RT studies, can be a good indication of language proficiency of multilinguals.

Spearman rank correlation coefficients were calculated to explore whether there is an association between participants’ RT and VS. The analysis suggests that in English there is
Results

213

Results

a significant weak negative correlation between the mean RTs and vocabulary scores: \( \rho = -0.29, p < .05 \). In Romanian \( \rho = -0.08, p > .05 \) and Hungarian \( \rho = -0.04, p > .05 \), however, despite the negative association, the correlations are non-significant. This suggests that the larger students’ vocabulary is, the faster they answer items in English, but this is not pertinent to the L1 and L2. Although, this was anticipated in Hungarian as the variance in the scores is minimal. A significant correlation also exists between mean RTs in English and Romanian \( (\rho = 0.63, p < .001) \), which can be interpreted as, the faster students are in their L2, the faster they tend to answer items in their L3. The associations between Hungarian RTs and the L2 and L3 respectively are much weaker. Between the L1 and L2 \( \rho = 0.48, p < .05 \) and between the L1 and L3 \( \rho = 0.21, p < .05 \), nevertheless still significant.

B. Speed of lexical access and the size effect

Of significance to both studies presented herein is that the academic year of study does not provide a clear distinction between proficiency levels. Therefore, relying on Laufer and Nation (2001) and Tanabe (2016), the participants were divided into two groups identically to RQ2 to establish whether more proficient students also showed faster accessibility in the tested FLs. Strictly speaking, this verifies whether students with a larger VS respond faster on form-meaning recognition tests than their lexically less proficient peers, termed size effect. The following table reports the descriptive statistics for the High and Low groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>ENVST</th>
<th>ROVST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test Score</td>
<td>RT</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>High</strong> ( (n = 28) )</td>
<td>47.68</td>
<td>4.99</td>
</tr>
<tr>
<td><strong>Low</strong> ( (n = 26) )</td>
<td>33.96</td>
<td>4.83</td>
</tr>
</tbody>
</table>

In response to RQ2, it has already been reported that the difference in VS between the High and Low groups is significant. Hence, only the difference in RTs are explored. As the Low group’s mean RT variable violates the assumption of normality, the Wilcoxon test is
utilised. According to this, students with a larger VS are significantly faster in English as the difference between the mean RTs is significant: \( W = 239, p < .05 \) with a medium effect size \( R^2 = .30 \). However, as the close means on the Romanian test suggest, there is no significant difference between the High and Low groups’ RTs (\( W = 332, p > .05, R^2 = .07 \)). The reason for this probably lies in the fact that the more proficient the group is in a language, the less variability can be detected in their RTs. This is actually further evinced by the fact that Hungarian RTs are basically identical (High group: \( M = 6.48 \); Low group \( M = 6.30 \)) if the Romanian proficiency split is used, for example, but the same tendency is observable with the English group split. Thus, regarding the size effect, it can be construed that in the case of multilinguals, the speed of lexical access is affected by proficiency, but apparently only in the least proficient language, which is the L3 (English) in this case. Theoretically, and tentatively, if any L4 or Ln were tested, this would have likely resulted in slower RTs as a function of decreasing proficiency.

C. Speed of lexical access and the effect of word frequency

Previous research suggests that RT is not just VS dependent, but it can also be described as a function of word frequency (Harrington, 2006; Laufer and Nation, 2001; Tanabe, 2016). The frequency effect in relation to RT postulates that lexical access becomes systematically slower as the frequency of items decreases.

Table 5.14 Mean RT distributed by frequency

<table>
<thead>
<tr>
<th>Test</th>
<th>Total</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENVST</td>
<td>M</td>
<td>12.52</td>
<td>10.26</td>
<td>11.52</td>
<td>12.44</td>
<td>13.57</td>
<td>17.73</td>
<td>13.34</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>6.96</td>
<td>5.52</td>
<td>6.64</td>
<td>7.69</td>
<td>7.37</td>
<td>7.27</td>
<td>6.82</td>
</tr>
<tr>
<td>ROVST</td>
<td>M</td>
<td>9.83</td>
<td>8.72</td>
<td>8.50</td>
<td>9.80</td>
<td>11.55</td>
<td>10.62</td>
<td>11.01</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>5.57</td>
<td>4.85</td>
<td>4.98</td>
<td>5.20</td>
<td>6.09</td>
<td>6.27</td>
<td>5.82</td>
</tr>
</tbody>
</table>

* Please note: mean RTs are total RT divided by the number of hits on each level (by-item) expressed in seconds. T1 corresponds to RTs on band one and two (first 2000 words, E1 and R1)

Table 5.14 illustrates that the mean RT is roughly inversely proportional to the item frequency. As items become less frequent, on average, it takes longer to answer them. This concurs with previous studies. Generally, this is the tendency on both tests. For instance, on
the ENVST compared to the first band (T1) where the average speed of response to one item is ten seconds, by and above the T4 band this gradually increases to around 14 seconds. Whereas on the Romanian test, the lowest average speed is nine seconds on T1 and T2, and it gradually increases to approximately ten seconds, except for T4 which stands out as having the highest time of 12 seconds.

Comparatively, in Figure 5.2 where on the same frequency band (R4) a slight dip in scores can also be noticed, this suggests that some of the items on this level might be more unfamiliar for some reason to students than on the surrounding bands. The item analysis conducted for RQ2 has already indicated that the item *monopol* attracted very few hits, which explains the dip on R4 in Figure 5.3. On the other hand, the RTs on the ten items on this level suggest that three items, two non-cognates (*moft* [caprice] and *abține* [refrain]) and one cognate (*păgân* [pagan]) have slightly higher RTs. This stipulates that words on T4 behave differently in terms of number of responses and mean RTs for different reasons. Nevertheless, considering the length of the test and the multiple-choice format, by and large, the assumption that RT proportionally increases with decreasing word frequency seems justified. To further substantiate this, mixed models were fitted using the RT variable as an outcome, word frequency as a predictor, and exploring whether this is affected by other factors.

The hierarchical modelling commenced with the English results. Compared to the baseline model, introducing item as a random intercept emerged as a significant model: $X^2(1) = 372.23$, $p < .001$. Following this, frequency was added as a fixed effect, which also emerged as a significant contributor to the model: $X^2(6) = 16.81$, $p < .05$. At this stage, the interest fell on whether there is variation between items and frequency, so frequency was added as a random slope. This, however, did not significantly change the model. Next, I wanted to explore whether the fixed effect of cognates varies across frequency levels, therefore, the binary variable of cognate and non-cognate was added as a random slope to
Chapter 5: Study Two (S2)

the block, which did not significantly improve the model and neither did cognacy as a fixed effect. The final model according to the R-squared values $R^2_M = .06$ and $R^2_C = .22$ only explains 6% of variance if only the fixed effects were used and 22% including the random effect. The final model is broken down in Table 5.15.

With regard to the Romanian RTs, in comparison to the null model, adding items as a random intercept, significantly improved the model: $X^2(1) = 437.41, p < .001$. In the next step, frequency was included as a main fixed effect, however, the change in the log-likelihood ratio was minimal: $X^2(6) = 10.92, p = .09$. Despite the introduction of frequency not majorly contributing to the model, for comparative purposes with the English one, it was retained in the model. In the ensuing, I added cognateness as a random effect, and as a fixed effect in turn. These additions did not emerge as significant improvements to the model. The final model thus explains 4% of variance considering only the fixed effects and 24% with the random effect ($R^2_M = .04$ and $R^2_C = .24$). The two final models are reported in the following table.

**Table 5.15 Mixed model of RT on the ENST and ROVST**

<table>
<thead>
<tr>
<th>ENVST</th>
<th>ROVST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Intercept</td>
<td>10.32</td>
</tr>
<tr>
<td>ET2</td>
<td>1.75</td>
</tr>
<tr>
<td>ET3</td>
<td>2.53</td>
</tr>
<tr>
<td>ET4</td>
<td>3.49</td>
</tr>
<tr>
<td>ET5</td>
<td>4.91</td>
</tr>
<tr>
<td>ET6</td>
<td>4.04</td>
</tr>
<tr>
<td>ET7</td>
<td>4.17</td>
</tr>
</tbody>
</table>

Sig. codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5.15 reports the RT models including frequency as a predictor. In English, it is evident that the positive $b$ values indicate that RTs increase systematically compared to high-frequency items (ET1). In Romanian, however, despite the positive association as evidenced by the positive $b$ values, the effect of frequency is not significant at all levels. Following this, exploring whether frequency is a significant predictor of L1 RTs would be in place.
However, similarly to the Hungarian scores, RTs are very consistent and due to this lack of variance, frequency does not appear to affect the model significantly. In other words, in the L1 the frequency of items does not influence the speed of lexical access.

D. **The effect of word cognateness on RT**

This section explores the facilitation effect of cognates in relation to RTs. Figure 5.13 below reports the cognate scores separated by multilateral and bilateral cognates (e.g. in English, the RO column reports the mean RT for the bilateral cognates, while the RO-HU reports the mean RT for the multilateral cognates). This differentiation is further investigated in the forthcoming. First, however, the difference between RT to cognates (bi- and multilateral combined) and non-cognates is explored.

![Figure 5.13 Mean RT divided by language and cognate status](image)

As a first step in exploring the difference in speed of lexical access to cognates and non-cognates, the RTs on the ENVST were analysed through a mixed-model analysis. Similarly to previous analyses, the RTs were the outcome variables, following which participants and proficiency are introduced as random effects. This resulted in a significant
contribution to the model: $X^2(3) = 417.25, p < .001$ and corroborates that RTs vary across participants and proficiency levels. Next, the cognate status of items was added as a fixed effect to the model. According to the maximum likelihood ratio, the addition of cognateness as a fixed effect did not significantly improve the model. Updating with a random slope for cognates did not contribute to the model either. Furthermore, the addition of proficiency as a fixed effect or the interaction between cognateness and proficiency, did not result in a significant change either. Therefore, it can be concluded that in English, cognate and non-cognate items were responded to at the same speed and this does not vary across proficiency groups. The following interaction graph illustrates this.

Figure 5.14 Interaction between proficiency and cognateness on the ENVST

Figure 5.14 illustrates that although there is a difference in the mean RTs between students with a large and small VS, cognates and non-cognates are responded to at the same speed by both groups. In the case of the Low Group, non-cognates appear to be responded to at a slower rate than cognates, however, as the mixed model indicates this difference does not emerge as significant.
The same procedure was followed with the RT on the ROVST. Compared to the null model, introducing the random effect for participants and proficiency significantly improved the model: $\chi^2(3) = 252.46, p < .001$. By updating with the fixed effect of cognates, the model significantly improved: $\chi^2(1) = 4.83, p < .05$, which means that Romanian cognate items are recognised faster than non-cognate items: $b = 0.46, t(2415) = 2.20, p < .05$. Finally, proficiency and the interaction between proficiency and cognateness was added to the model, however, neither the higher order interaction nor the main effect of proficiency contributed significantly to the model. In other words, lexical access for cognates does not vary across proficiency levels.

As the multilevel analysis regarding the Romanian RT and Figure 5.15 stipulates, cognates on average are recognised significantly faster than non-cognate items, however, despite the tendency for the Low group to have slightly higher means, there is no significant variation across proficiency. This suggests that despite Romanian item recognition in terms of speed are influenced by cognateness, the VS of students does not influence this.
With regard to the L1 RTs, the mixed model indicates that adding a random intercept for participants significantly improves the model compared to the baseline model: $\chi^2(1) = 205.82, p < .001$, however, adding either proficiency or cognate status of words as random slopes, in turn, leads to no improvement in the model. Introducing the fixed effect of cognates shows a significant contribution to the model: $\chi^2(1) = 5.61, p < .05$. This demonstrates that responses to non-cognate items were significantly slower than to cognate items, $b = 0.26, t(3522) = 2.37, p < .05$.

![Figure 5.16 Interaction between proficiency and cognateness on the HUVST](image)

Figure 5.16 Interaction between proficiency and cognateness on the HUVST

Figure 5.16 is a confirmatory visualisation of the aforementioned model. There is no effect of proficiency, students with larger and smaller VSs respond at roughly the same rate to cognate and non-cognate items. Nonetheless, it demonstrates that Hungarian cognates have lower RTs than non-cognates.

Altogether, it appears that the cognate status of items does not appear to influence RTs in the least proficient language under study, English L3 in this case, whereas in the more proficient languages, the L1 and L2, cognates tend to be answered on average faster than
Results

non-cognates. Importantly, in the case of the group with lower lexical proficiency, there seems to be a tendency for answering cognates at a faster rate in English, albeit, this did not lead to significance. Overall, it can also be concluded that RTs are only affected by proficiency in English, whilst in the other languages, RT appears to be constant across the proficiency groups.

Previously in this section, it has been mentioned that cognates at the design stage were categorised into bi- and multilateral cognates (see Chapter 3). This contrast has been illustrated already in Figure 5.13. The hypothesis behind this is that compared to bilateral cognates, multilateral cognates, if co-activated in all three languages (non-selective access view) elicit a response latency compared to bilateral cognates due to the increased cognitive load. Should this difference be non-significant, it would be conceivable that cognates are accessed selectively, or separately, in different languages, which in turn would mean cognates on multiple-choice tests are recognised faster than non-cognates in the more proficient languages (L1, L2) and this is not affected by the type of cognate (bi- or multilateral). Conversely, it is also conceivable that due to the multiple-choice test, the non-contextual sentence, perhaps frequency, and/or other latent factors, the amount of time spent on each item becomes less susceptible to the cognate status of items. In other words, the VST test format with an employed temporal measure might not be sensitive enough to distinguish RTs between different types of cognates as other factors might also be at play.

To explore this hypothesis, the above models have been reconstructed with orthogonal contrasts. These compare in contrast one: non-cognate items to cognate items; and in contrast two: bilateral to multilateral cognates only. The first contrast has been discussed above, so I proceed to the second contrast. In all three languages, the bilateral versus multilateral cognate contrast emerged as a non-significant predictor, which means that bi- and multilateral cognates were recognised on average at the same speed (see Figure 5.13). Whether this is an artefact of the type of test and/or the implementation of the temporal
measure, or it points towards selective access has to be subjected to further explorations in future research. Tentatively though, this result can be connected to the Bartolotti and Marian (2017) study, which served as indication that wordlikeness (orthographic probability) in one background language is enough for a facilitative effect to emerge.

E. The effect of other word characteristics on RT and VS

This section examines other factors that can influence word difficulty in terms of speed and accuracy. Such factors herein were POS and word length, operationalised as the number of syllables. On the one hand, the aim was to conduct a more controlled online study in which not just the frequency effect is investigated, but other lexical characteristics as well. On the other hand, this would have allowed directly comparing the results with studies, such as Daller and Milton (2007) or Willis and Ohashi (2012) on two distinct, but inherently connected dimensions, RT and accuracy. However, neither in the case of RT nor in the case of accuracy have either word length or POS emerge as significant descriptors of models, so the decision was to drop these variables from the previous analyses. To ensure consistency, though, the mean RTs are reported in the languages under study based on these variables.

Table 5.16 Mean RT based on the different item properties

<table>
<thead>
<tr>
<th>Property</th>
<th>ENVST</th>
<th>ROVST</th>
<th>HUVST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Verb</td>
<td>12</td>
<td>12.27</td>
<td>7.18</td>
</tr>
<tr>
<td>Adjective</td>
<td>9</td>
<td>10.95</td>
<td>6.63</td>
</tr>
<tr>
<td>Noun</td>
<td>49</td>
<td>12.85</td>
<td>6.94</td>
</tr>
<tr>
<td>Bilateral C</td>
<td>19</td>
<td>12.39</td>
<td>6.74</td>
</tr>
<tr>
<td>Multilateral C</td>
<td>20</td>
<td>12.47</td>
<td>6.52</td>
</tr>
<tr>
<td>Non-cognate</td>
<td>31</td>
<td>12.64</td>
<td>7.42</td>
</tr>
<tr>
<td>Syllable no: 1</td>
<td>19</td>
<td>12.31</td>
<td>7.02</td>
</tr>
<tr>
<td>Syllable no: 3</td>
<td>12</td>
<td>12.56</td>
<td>6.93</td>
</tr>
<tr>
<td>Syllable no: 4</td>
<td>7</td>
<td>15.12</td>
<td>6.37</td>
</tr>
<tr>
<td>Syllable no: 5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Syllable no: 6</td>
<td>1</td>
<td>14.22</td>
<td>6.05</td>
</tr>
<tr>
<td>TOTAL</td>
<td>70</td>
<td>12.52</td>
<td>6.96</td>
</tr>
</tbody>
</table>

Table 5.16 confirms that, maybe, due to the number of items and test format, the properties
that have been controlled for had no effect on the speed of responses. It demonstrates that
despite the word length of items, the mean RTs stay relatively constant across languages.
Similarly, POS does not seem to influence majorly the mean RTs either. Cognateness has
already been investigated in the above.

Table 5.17 Mean item facility based on the different test properties

<table>
<thead>
<tr>
<th>Property</th>
<th>ENVST</th>
<th>ROVST</th>
<th>HUVST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Verb</td>
<td>12</td>
<td>54.00</td>
<td>25.19</td>
</tr>
<tr>
<td>Adjective</td>
<td>9</td>
<td>55.67</td>
<td>25.63</td>
</tr>
<tr>
<td>Noun</td>
<td>49</td>
<td>60.35</td>
<td>24.52</td>
</tr>
<tr>
<td>Bilateral C</td>
<td>19</td>
<td>52.89</td>
<td>24.14</td>
</tr>
<tr>
<td>Multilateral C</td>
<td>20</td>
<td>68.35</td>
<td>25.14</td>
</tr>
<tr>
<td>Non-cognate</td>
<td>31</td>
<td>55.94</td>
<td>23.32</td>
</tr>
<tr>
<td>Syllable no: 1</td>
<td>19</td>
<td>56.42</td>
<td>28.54</td>
</tr>
<tr>
<td>Syllable no: 2</td>
<td>31</td>
<td>60.00</td>
<td>21.54</td>
</tr>
<tr>
<td>Syllable no: 3</td>
<td>12</td>
<td>63.25</td>
<td>25.78</td>
</tr>
<tr>
<td>Syllable no: 4</td>
<td>7</td>
<td>47.43</td>
<td>25.18</td>
</tr>
<tr>
<td>Syllable no: 5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Syllable no: 6</td>
<td>1</td>
<td>83.00</td>
<td>NA</td>
</tr>
<tr>
<td>TOTAL</td>
<td>70</td>
<td>58.66</td>
<td>24.55</td>
</tr>
</tbody>
</table>

Table 5.17 illustrates that multilateral cognates have been answered on average by more
students than non-cognate and bilateral cognates on both the ENVST and ROVST. However,
the difference is not large enough to emerge as significant. In summary, it is entirely
conceivable that in the case of proficient multilinguals on a multiple-choice test where a non-
contextual sentence is also provided, RTs and accuracy are not influenced, given the group
sizes, to the same extent as with lower proficiency Japanese L1 learners as in Willis and
Ohashi (2012). The results, therefore, seem to echo Daller and Milton (2007) who only found
frequency as a significant contributor to the model in the case of French L2 learners with
English L1. This is further elaborated on in the ensuing chapter.

In answer to RQ3, there are several important aspects to consider. First, it has been
corroborated that RTs align with lexical proficiency. Students’ speed of lexical access was
the lowest and most consistent in their L1, ensued by the L2 and the largest latencies were
shown in the L3. With regards to English, it has also been shown that faster students also
attained higher scores, but this was not the case in either of the other languages tested. Importantly, there also appears to be an association between the speed of lexical access in the L2 and L3 especially, albeit moderate linearity also exists between the L1 and L2/L3 as well. The second part of RQ3 addressed whether subjects with a larger VS display a tendency towards lower accessibility rates than their lexically less able peers. The non-parametrical tests employed unveil that in the least proficient language, English as it were, lexical access appears to be faster in the case of larger VS. This difference, however, was not significant in the L2.

A quintessential issue to lexical testing that has been explored throughout the thesis is the effect of cognates. It appears that in English, the more proficient group recognises cognates and non-cognates faster than students with a lower VS, however, the difference is not statistically significant either for proficiency or cognateness. Thus, from an RT point of view, the evidence suggests that proficiency and cognate status of words do not affect lexical access. With respect to the Romanian RTs, the analysis indicated that cognates are answered quicker than non-cognates, but the difference between high and low groups does not reach statistical significance. This finding has been confirmed by RTs on the Hungarian test as well since non-cognates displayed longer latencies than cognate items. This difference in the L1 and L2 gives way to interpret the findings relating to English RT as inconclusive regarding the cognate status of items.

On the one hand, it is amenable to theorisation that due to the lower proficiency, the cognateness of items has a limited influence on lexical access. On the other hand, the lack of a facilitative effect of cognates in relation to RT could be due to the fact that students do not necessarily regard cognates as affordances and might be sceptical about their meanings. Conversely, the finding could be an artefact of the methodology as the four options and the non-defining sentence could yield longer cognitive processing latencies, which might not be directly or exclusively influenced by the cognate status of words. These pertinent issues will
be revisited in the next chapter.

The last major issue in relation to RQ3 has been the effect of frequency as a function of RT. Generally, both the L2 and L3 tests manifested a roughly inversely proportional relationship between RT and item frequency. Items of higher frequency show a tendency towards faster lexical access, however, this tendency was empirically substantiated on the ENVST only. In the more proficient L2, it appears that RT is less affected by the frequency of items.

5.8.5 The relationship between test scores and learner-related factors

Corresponding to S1, the purpose of the questionnaire (see Appendix 7) was to add further validity to the empirical data explored hitherto and to aid the triangulation of specific aspects that are central to multilingualism. To this end, the first question addressed the age of onset of learning the L2 and L3. Based on the data in S1, it has been extrapolated that there was a significant negative correlation between the age of onset of learning both languages and the test scores. In relation to the present study, the Spearman rank correlations stipulate that there is a weak negative correlation between the ROVST scores and the age participants started to learn Romanian: $\rho = -.27, p < .05$ ($N = 54$ throughout this section), albeit, the correlation is non-significant in English. The reason for this is that regardless of the age they started learning English, students might have had different amounts of exposure. For instance, some might have been in intensive English classes (6-8 hours/week) in school or taken private lessons to prepare for university. Retrospectively, if there is no obvious association between age of acquisition and L3 proficiency, this makes employing the variable to check for homogeneity in the population (section 5.8.2) questionable at best. However, it remains the case that the GPA variable is still a valid indicator of the homogeneity of the population.

Additionally, the correlations confirm that the higher students’ English VS is: the more proficient they feel in that language ($\rho = .50; p < .001$ unless otherwise specified); the
more confident they feel using it in everyday situations ($\rho = .52$) and in formal settings ($\rho = .52$). Similarly, the analysis confirms that there is a moderate correlation between the ENVST scores and students’ self-rated confidence levels in using English in a work situation ($\rho = .47$), and if they moved to an English speaking country ($\rho = .41$).

Considering the ROVST scores and the self-ratings, the results indicate that the test scores have a significant relationship with: self-rated proficiency ($\rho = .50$), confidence levels in everyday ($\rho = .53$) and formal ($\rho = .50$) situations. These results show that the VSTs appear to be relatively good indicators of how proficient and confident students feel in using a FL, at least in this cultural context (e.g. Japanese students tend to be very modest; Kamimoto, personal communication). Associations, or correlations, of this nature do not, of course, implicate a direct or causal link, albeit this may be conceivable. Learners might have a clear idea of the academic grades or levels of communicability they aspire to achieve in different settings and having the appropriate lexical knowledge can support them to expedite their acquisition journey. This, in turn, has the potential to raise their self-perceived proficiency and confidence levels.

Table 5.18 summarises the responses in percentages to proficiency and confidence questions from the questionnaire. This reveals that only 11% of the students consider themselves not proficient in English, compared to 30% in Romanian. Taking into account the fact that all students have lived most of their lives in Romania, this is a somewhat surprising finding. However, it is in line with findings in S1. Regarding confidence ratings in everyday situations (e.g. in a restaurant) and formal settings (e.g. job interview) it is suggested that 39% and 54% students respectively do not feel confident enough in using their L2, despite the higher scores on the ROVST than in English. In comparison, only around 25% of the students do not feel confident in English and the majority rated themselves as fairly to very confident in both situations. 74% of the students claim that they could use English efficiently at an international company, compared to 78% when it comes
Questions 7, 8, and 9 had the aim of gathering information on how students perceive the links between their languages and what strategies they use. According to the answers, when browsing (in their ML) for an English word while writing, 83% of the students think of suitable Hungarian words and translate those, instead of relying on their Romanian knowledge. This finding, besides confirming the findings in S1, reflects the fact that most of the students have Hungarian NS English teachers in schools and thus the link between these two languages becomes stronger and, arguably, with a faster or more direct lexical access connection within the network. Simultaneously, when reading, 48% of the students translates unknown L3 words into Hungarian and 46% tries to guess from context. 6% chose the ‘other’ strategy option and no student chose options b (‘I try to think if I’ve seen the word in Romanian’) or c (‘I try to find a Romanian translation’). This clearly shows that the L1 to/from L3 is a very strong link, despite the typological similarities between the L2 and L3. The last question (9) reveals that, contrary to students’ strategies while reading or writing in English, over half of them (54%) think that Romanian is useful and only 11% consider it to be confusing or misleading.

Question 10 provided students with the opportunity to comment on the cognate-awareness tasks and express their opinion on how much they think Romanian would or would have helped them in learning English. The large majority of students considered the

### Table 5.18 Descriptive statistics for self-rating questions (N = 54)

<table>
<thead>
<tr>
<th>Proficiency self-rating</th>
<th></th>
<th>Confidence in everyday situations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Romanian</td>
<td>English</td>
</tr>
<tr>
<td>Rating</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>%</td>
<td>4</td>
<td>7</td>
<td>41</td>
</tr>
<tr>
<td>Confidence in formal settings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>%</td>
<td>6</td>
<td>19</td>
<td>50</td>
</tr>
</tbody>
</table>

*Note: 1 is ‘not very’, 3 is ‘fairly’, and 5 is ‘very’ proficient/confident, the second row (%) provides the percentage of students who chose that rating (five-point Likert-scale)
EI on cognates and the task useful. Their comments point to the fact that learning their L3 is definitely enhanced by Romanian. Some of them also emphasised that this is probably the reason why Romanian NSs are better in English than Hungarians (confirmed by Molnar, 2008). Due to the cognate facilitation effect, many of the participants admit that learning English through Romanian would have been easier, but it also depends on the level of the L2. Some of the following examples from students’ comments reflect these issues:

a. First of all I liked the text, since it is about Madagascar. (Yupiii) Making the comparison between the 3 texts I came to the conclusion that these languages share much more than I have imagined before. When I make translations I never think in Romanian however now it seems it is a good idea. […] On the other hand I believe, Hungarians cannot speak Romanian on a level to learn another language on its basis […]. SS50

b. It would be very useful because having similar base or spelling with Romanian words would lead to a logical acquisition of English vocabulary and it would make it easier to retain the words and use them correctly. SS25

In essence, the analysis of the questionnaire seems to indicate that there is no discrepancy between the test results and how students perceive their proficiency and confidence levels in English and Romanian. Furthermore, in contrast with the proximity between the two additional languages and the fact that most students accept that Romanian can be helpful when dealing with (learning, processing, and using) English, their strategies stipulate that simply making L2 and L3 connections is unnatural, which in turn, suggests that the study has teaching and learning implications.

Finally, in conclusion to this chapter, the major findings that have been revealed appear to confirm, expand on, and refine the results of S1. Furthermore, Milton, Alexiou, and Mattheoudakis relying on their research into the recognition of aural and written word recognition arrived at the conclusion that ‘the form of a test of vocabulary knowledge, especially size, may be crucial if the desire is to capture the full extent of the learners’ foreign language lexicon’ (2014: 28). The empirical studies in this thesis extend this assertion by tentatively claiming that the aspiration of understanding the multilingual lexicon, and by
extension vocabulary acquisition, may be expedited if learners’ background languages are accounted for. The form-meaning mapping knowledge appears to be particularly relevant and revealing, as prior lexical knowledge can be a source for establishing lexical-semantic links in the case of interlingual homographs that share a common meaning. It also emerges that while frequency of word occurrence is quintessential for such research, it only provides a partial picture and accounting for cognate knowledge can be a meaningful addition as it can be assessed comparatively objectively using similar methodology. The next chapter aggregates the themes and findings from the empirical studies into a discussion and seeks to create a link with prior research, generalise the findings, point out important implications, and limitations.
Chapter 6: Discussion

The introduction set out two broad aims. On a theoretical level, this thesis aimed at exploring potentially influential factors regarding the vocabulary knowledge of multilinguals and, within this, investigate the relationship between learners’ interlanguages in a learner-centred framework. On a more substantive level it proposed to challenge the idea whether frequency can solely account for the mechanisms of LL or the size of the ML as a resource for language processing. It emerges that in the case of the typologically close languages under study herein, there is a strong link between learners’ background languages, and Romanian can serve as a resource for vocabulary recognition in English. This is in full accord with and expands Fitzpatrick and Milton’s conclusion: ‘studies have shown time and time again that learner performance on one dimension of word knowledge enables us to predict performance on other dimensions’ (2014: 177). The two empirical studies presented take this a step further by demonstrating that word knowledge in one language is closely interlinked with other languages.

Consequently, it transpires that in order to understand the learning process of lexis, it is essential to uncover what is already part of learners’ vocabulary, and what other factors influence lexical acquisition. The present and concluding chapters, therefore, further elaborate on the emergent findings by relating them to previous studies, emphasising the magnitude of the findings, and also acknowledging the limitations and implications for future research in the area. For reasons of systematicity, the sections chiefly follow the order of the RQs. One exception to this is section 6.7 which adds to the evaluation and discussion of the results of a preliminary follow-up research on the distribution and proportion of Romanian cognates in the English language.
The effect of explicit instruction on cognates (RQ1)

6.1 The effect of explicit instruction on cognates (RQ1)

The studies reported in this thesis do not claim to address all aspects of vocabulary knowledge and measuring VS in the case of multilinguals. However, through these empirical explorations I have been able to expose and problematise some pertinent issues to vocabulary testing and, by extension, lexical acquisition. First, the experimental paradigm is discussed.

RQ1 explored whether EI on cognates, i.e. pointing out structural similarities between Romanian and English interlingual homographs with a semantic overlap, would bear an effect on participants’ VS. Although the response to this question dissented from the initial expectations, anecdotally speaking, it is often said that in educational research non-significant findings can also result in invaluable extrapolations. Some of these are elaborated on in the following.

Although neither of the studies were designed to directly explore cognate recognition cross-linguistically (i.e. employing identical items in the tests), it has been hypothesised that if the item is known in Romanian, then a cognate awareness-raising exercise should facilitate cognate recognition in English. This is premised on the idea that cognates in Romanian, as Latinate words, may be of higher frequency than in English, and therefore, these should be conceptually more familiar to learners (e.g. Lubliner and Hiebert, 2011; Nagy et al., 1993). On the other hand, as participants live in Romania, they should have a higher exposure both personally and educationally to Romanian and be, thus, more proficient in it. This hypothesis has been falsified in relation to adult and proficient learners by both S1 and S2, thus confirming Molnar (2010).

Molnar (2010) revealed that statistically there was no effect of the experimental treatment. The studies featured in this thesis fully confirm these results, regardless of the test used or the language tested (Molnar only measured VS in English). While she seems to ignore the statistical perspective and concludes that the slightly higher mean for her SG is evidence for the benefit of such instruction, I regard it comparatively as a more multi-faceted
issue and the results are discussed in contradistinction to her findings. The basis for this is principally twofold.

The first argument is predominantly rooted in the selection of participants. Both studies involved proficient and cognitively developed language learners from high-prestige secondary schools, in Molnar (2010), or undergraduates in the present case. It can be argued therefore that participants are competent language users with not just recognisable LL experience, but advanced metalinguistic skills as well. Thus, the non-significant difference between the SG and CG potentially can be taken as evidence for the Parasitic Model (e.g. Hall, 2002; Hall and Ecke, 2003; Hall et al., 2009).

According to this model, new items are acquired either on lexical or conceptual level through similarity detection and exploitation, or pattern-matching and assimilation. Consequently, cognates are recognised intuitively and once a representation in one of the background languages exists, a link is formed to this, serving as a bridge to the new item (e.g. Hall et al., 2009). Bartolotti and Marian (2017) extend this and posit that lexical items benefit from having connections to at least one language. Due to the robust correlation between cognate scores in the two languages, the facilitative effect extrapolated herein, and the English-Romanian typological proximity on a lexical level leads one to believe that cognate recognition is automatic indeed, hence the lack of effect of the EI.

There is reasonable evidence to speculate though that this result also pertains to the combined effects of language proficiency and age. García (1991) for instance intimated that young learners could benefit from cognate instruction. Proctor and Mo also report on the benefits of ‘Al, a bilingual literacy coach embedded in the environment’ who provided Spanish-English cognate examples to bilingual young people, following which they matched NSs on a cognate awareness task (2009: 129).

Contiguously to these cognate benefits, Jessner (1999), Otwinowska (2016), and Ringbom (2007) iterate that less advanced learners require training so as to maximise their
chances of capitalising on cognate or alternative inter-lingual congruencies. Furthermore, Hancin-Bhatt and Nagy (1994) and Nagy et al. (1993) infer that as learners become older, their cumulative LL experience and metalinguistic awareness allows them to profit more from cognate knowledge, thus, cognate recognition is not just proficiency, but age dependent as well. With regards to languages with non-congruent scripts, however, Helms-Park and Perhan (2016), found that learners can benefit from EI, which can suggest that in the case of typologically more distant languages, even short cognate-awareness treatments can result in substantial pedagogical implications.

Consequently, it seems rational to conclude that the results of the current studies on the one hand, tentatively support the Parasitic Hypothesis and are in full accord with Zipf’s (1949) least effort principle and other economic reasons that might be governing the underlying factors of mental representation, processing, and use of lexis (Hall, 2000). However, this is based on the similarities between English and Romance languages and, adult, proficient learners. Antithetically, these studies cannot confirm whether the automatic recognition of a cognate, provided that it is known in any of the background languages, will stand in the case of different language constellations, especially with non-congruent scripts, or with beginner learners across different age groups. Thus, further studies, preferably longitudinal, are required to explore the intrinsic nature of cognate recognition that involve a wide range of languages and age groups, and extend beyond the written and receptive modalities.

The second argument is majorly grounded in the choice of tests. Both test formats employ frequency-sampled items, thus, although, it can be assumed that the tested words will more or less be representative of other words with similar frequencies, whether the test format is capable of detecting the outcomes of a short awareness-raising intervention is questionable. Since the target items featured in the EI are randomly selected items, which do not form part of either test versions in any of the tested languages, it may be extrapolated
that the intervention does not lead directly to vocabulary gains per se. In other words, the tests, in actuality, measure meaning-based receptive knowledge and not how much students know about these words, depth of knowledge as it were, or how easily or fluently they can employ or transfer them to the alternative language. Therefore, not knowing the actual meaning of some cognates in the tests is unlikely to be influenced by the instruction.

As a cognate recognition strategy, however, the instruction served beneficial in demonstrating that taking into account the learners’ proficiency, age, and language background, the search for form-meaning in the two languages is apparently automatic, a strategy that students use intuitively. Future studies could explore this via translation exercises or elicitation tasks in which, were the cognate words provided in the alternative language, regardless of the instruction, this would offer additional evidence in support of the Parasitic Model.

**6.2 The relationship between the L2 and L3 lexicons (RQ2)**

In answering RQ2, the findings have been partially unanimous and in the affirmative. These confirm that overall there is a strong, sizable relationship between the background languages and evidence for this comes from various sources, which are discussed in this and the following subsections.

It is apparent from both studies that students’ VS in the two languages is interdependent as indicated by the strong correlations between overall and cognate scores in the L2 and L3. This is entrenched in Cummins’ Developmental Interdependence Hypothesis (e.g. 1991), which can be extended to the relationship between the L2 and further languages (see Cenoz, 2003), and postulates that skills can be transferred from the L1 or L2 to the L3. The assumption here is premised on the fact that since students develop certain skills and knowledge in their Romanian L2, during the acquisition of English they can rely on such competencies and linguistic resources, and this will facilitate LL. In the studies presented in this thesis, this focal aspect was receptive lexical knowledge. Specifically, the analyses
explored VS in the L2 and L3 through recognition tests that measure vocabulary knowledge, based on the form-meaning mapping skill in a cognate-non-cognate paradigm.

The analysis in S1 corroborates that the relative VS scores explain 62% of the variance. Concurrently, the VSs in S2 also indicate a strong relationship between lexical proficiency in the two languages and the test scores account for 35% of the variance. These medium to large effect sizes elicited by the tests are in line with Molnar’s (2010) findings. Despite the fact that she relied on subjective teachers’ ratings, her findings infer a medium effect size or a variance of 17% accounted for by the data. A transformation of correlation coefficients into z-scores (the calculator used can be found on: http://www.danielsoper.com) to determine whether there is a significant difference between Molnar’s and the two studies presented in the thesis indicates that the difference is significant in the case of S1, which utilised the same test as Molnar. However, the difference is non-significant in the case of S2.

The reasons for this are potentially twofold. Both Molnar and S1 found an increase in scores on the UWL. In S1 the correlation between the UWL and the 6k band was the highest due to the high number of cognate items. From one perspective, this emphasises the strong relationship between the two languages in terms of cognate knowledge. From an alternative perspective, the fact that the number of cross-linguistically similar items influences the correlation between the two levels suggests that a large proportion of cognates on one level can be problematic. Moreover, the results also indicate that cognate knowledge is a strong predictor between proficiency and overall lexical dexterity in the two languages.

The other reason for the lack of a statistically significant difference between Molnar’s study and S2, based on the z-scores, could be related to the VSTs employed. The reduced number of items might compromise the tests’ efficacy in estimating ‘exact’ VSs, despite that they better represent vocabulary knowledge on different frequency levels. Thus, it can be speculated that should the 140 items have been maintained, this would have resulted in a more robust correlation between the L2 and L3, and ultimately, a significant difference.
between Molnar’s study and S2.

Nevertheless, what is perhaps more important than the difference between the correlations in the three studies, is that they take the same directionality in confirming a strong relationship between English and Romanian lexical proficiency and in contributing further evidence to the Interdependence Hypothesis. Moreover, the larger effect sizes in the current studies suggest that measuring VS empirically through the Romanian lexical tests brings forth substantial evidence on the interdependence of lexical proficiency in the two languages and allows a close L2-L3 comparison in a meaningful way.

Regardless of the fact how one conceptualises the multilingual ML, whether it is one system that integrates all languages or separate networks linked to one conceptual system, the findings predominantly point into the direction that from a VS perspective, these (sub)systems within the ML are strongly interlinked. If the conservative and rather simplified assumption that cognates are all shared in a combinatorial fashion between the languages is taken, then, the data should suggest that cognate recognition in the two languages yields a robust relationship. This obtains in both S1 and S2.

Confirming Molnar’s findings further, which indicated a larger effect size for L3 cognate knowledge and Romanian proficiency, L2 cognate scores account for 76% of variance of L3 cognate scores in S1 and 40% in S2, compared to non-cognate scores, which explain 42% (S1) and only 10% (S2) of the variance. Although there is no evidence provided in Molnar (2010) regarding non-cognates and how they correlate with Romanian proficiency ratings, the above brings forth substantial evidence that cognates are strongly linked within the two languages. This would be further evidenced if the two studies allowed inferring that cognates known in one language are also recognised automatically in the other. Since both studies were designed to feature unique cognates in the L2 and L3, the data obtained cannot be interpreted this way, which can be also regarded as a limitation of the methodology and hence is further elaborated on in a forthcoming section.
Despite this, cognate recognition as a skill can also be linked back to the Interdependence Hypothesis. From the results, it can be construed that if cognates are recognised in the L2 to a certain extent, this should also pertain to the L3 as well. This is conferred by the large effect sizes regarding the cognate scores. This idea can also be taken further in the way that if students have highly developed cognate recognition skills in the L2 and L3, this should indicate not just overall lexical proficiency in the same language, but the other language as well. In fact, both studies demonstrate this tendency.

Specifically, S1 indicated that cognate scores in English explain 71% of the variance in the overall Romanian scores and 38% in S2. Parallel to this, the Romanian cognate scores account for 69% of the variance in the overall English scores in S1 and 41% in S2. Again, the difference between the two studies can be subscribed to the lack of academic words in S2 and, maybe, other latent factors. Nonetheless, these strong relationships on the level of cognates serve as further evidence that skills, particularly cognate knowledge, in one language are a strong predictor of overall VS in the other language. Although the two tests only allow comparing lexical knowledge overall and cognate or non-cognate recognition in the two languages, the two studies also investigated the relationship between lexical knowledge and overall proficiency in terms of academic achievement. As this adds further evidence to the interdependence between the languages, I now turn to discuss this briefly.

The studies presented in this thesis converge with previous findings corroborating that lexical dexterity is a relatively good indicator of academic achievement (e.g. Daller and Wang, 2016; Milton and Treffers-Daller, 2013). Milton and Treffers-Daller (2013) found a modest correlation between the vocabulary scores and average module marks. The highest explained variance was 23%. Taking into account that they employed a yes/no test, as opposed to a form-meaning recognition test, and that vocabulary knowledge is just one of the wide range of skills and knowledge that students need to survive in academia, it is a substantial finding.
Congruent with this is the finding in Daller and Wang (2016, but see also Daller & Xue, 2009; Daller & Phelan, 2013; Laufer and Goldstein, 2004; Yixin & Daller, 2015) in relation to L2 learners of English. These studies used different measures of vocabulary dexterity, yes/no tests or C-tests for instance, and found a significant correlation between lexical knowledge and students’ academic achievement.

The two studies featured herein are congruent with these findings to the extent that in S1 the English vocabulary scores account for around 40% of variance of the two academic indicators, and although to a lesser degree, the Romanian scores explain around 20% variance. Similarly in S2, English scores explain around 20% variance, albeit the Romanian VS indicate only a weak correlation with approximately 10% of the variance accounted for. This can be interpreted as further evidence for the Developmental Interdependence Hypothesis as, if vocabulary scores in English and Romanian were completely unrelated, Romanian vocabulary knowledge, which is not directly influential of English GPAs, would not result in a correlation with the academic performance indicators. However, since a relationship does exist, it can be assumed that academic achievement in an English degree, to a certain extent, can be explained by vocabulary knowledge in not just the L3, but the L2 as well, due to the interconnection between these languages. Additionally, taking the position that GPA should be an amalgamation of other linguistic knowledge and competencies, including metalinguistic and communicative competence, the results specify that lexical proficiency may be regarded as a key element of not just academic success, but the development of other abilities as well.

Another salient result that the two studies revealed, and slightly diverge in terms of postulations, also relates to participants’ proficiency in the two languages. S1 claimed that despite the fact that testees reside in Romania, start learning Romanian at a younger age with a higher exposure both formally, and presumably, informally as well, since they live in the target country, their L2 did not yield higher scores. Conversely, this hypothesis in S2 has
been partially borne out as Romanian lexical knowledge emerged as significantly higher than English L3. Since Molnar’s study did not empirically investigate Romanian VS and teachers’ ratings were not reported, a comparison between the two researches is impossible in this sense. Nevertheless, the diverging results pertaining to the two studies can be explained in three ways.

One interpretation to assume is that since the two studies feature different participants, in S1, the Romanian knowledge of participants was overall lower. This premise is also complemented by the questionnaire, as it indicates that students would feel more confident in using English than Romanian.

The other interpretation is that this is an artefact of the difference between the VLT and VST. The UWL, in the case of S1, inflated the scores primarily because of cognates and English cognates are more common words in a Romance language, thus increasing the probability of hits. This is further supported by Molnar’s findings as the scores on this level stood out there as well. Additionally, being undergraduates also increases their familiarity with these items. By contrast, as the ENVST and ROVST comprise items that are sampled from frequency levels only, no academic words are included (specifically from the UWL/AWL that is), this precludes the inflation of scores on any of the levels.

The third perspective relates to the word lists used. While both English tests deploy word families as a counting unit, the RWL is based on lemmas. Since word families considerably reduce the number of different basewords in a list compared to lemmas, it is quite conceivable that the Romanian tests underestimate vocabulary knowledge, albeit the exact effects of using lemmas in lexical studies are yet to be empirically substantiated (e.g. Kremmel, 2016; Milton, 2009).

In short, it appears reasonable to assert that despite the higher vocabulary score expectations in Romanian, due to the compilation of the VLT and participants’ language background, S1 failed to establish language dominance. This brings further evidence for the
inference that testees’ cultural, especially linguistic, background should be taken into account and an appropriate lexical test should be culturally independent to adhere to assumptions of validity.

This part of the discussion in relation to RQ2 mainly related to the lexicon of learners overall and other learner-related aspects that pertain to their lexical knowledge. The upcoming sections address RQ2 from a cognate perspective.

6.3 Lexical proficiency and cognate recognition

The empirical findings corroborated an additional issue that pertains to proficiency and cognate recognition. There is strong evidence suggesting that cognate recognition is proficiency dependent, which concurs with previous studies. Dressler (2000) and Otwinowska (2016), inter alia, proclaimed that proficient multilinguals who had a better understanding of cross-linguistic similarities, due to their cumulative LL experience, performed better on cognate items and showed increased awareness of cognates. Similarly, Hancin-Bhatt and Nagy (1994) or Méndez Pérez et al. (2010) also conclude that as more proficient learners have increased metalinguistic knowledge, they benefit more from the presence of cognates.

By contrast, Van Hell and Dijkstra (2002) claim that learners have to reach a certain threshold for the cognate facilitation effect to take place and, additionally, the effect is smaller in the L1. Others’ contention is based on findings indicating that cognate recognition is more noticeable on lower proficiency levels (e.g. Laufer and McLean, 2016; Elgort, 2013; Serrander, 2011).

The extrapolations from the analyses convey evidence for the former and latter findings simultaneously. Both studies indicate that more proficient learners also recognise more cognates, not just in their L2, but the L3 as well. In spite of this, S2 indicated that on the highest frequency level, students in the Low group match the scores of the High group. Similarly, while on low-frequency bands, the difference between the proficiency groups is
minimal, on mid-frequency items lexically more able students could rely more on the facilitative effect of cognates. This is so especially on the ENVST.

Regarding Romanian, on three different levels the Low group actually were on par with the High group’s the scores, which indicates that they benefitted from the facilitatory effect of cognates. This, and the fact that in Hungarian there was no cognate effect, tentatively supports Van Hell and Dijkstra (2002).

As a general conclusion to the relationship between proficiency and cognate recognition, it can be inferred that on more frequent items, the difference between proficiency groups diminishes. Thus, cognates probably facilitate lexical recognition for participants with a smaller VS, which echoes the extrapolations by Laufer and McLean (2016) and Elgort (2013). However, as a general trend, the more proficient group recognises considerably more cognates, and non-cognates, on the alternative bands, thus supporting previous findings suggesting LL experience and metalinguistic knowledge aid cognate recognition (e.g. Hancin-Bhatt and Nagy, 1994).

6.4 The effect of cognates on testing

Hitherto, the discussion largely explored the L2-L3 interrelation. Concomitant with this issue is language proximity as well, and it needs to be noted that probably a further explanation of the strong relationship between the above-described proficiency in the L2 and L3 is prevalent to the close lexical relationship between English and Romanian. In Ringbom’s intimation (2007) FL acquisition is principally attaching or associating the new, or unknown, with the already known. Therefore, as both studies reveal, overlapping vocabulary between the two languages can serve as a catalyst for word recognition and, in turn, potentially accelerate learning. By way of extension, it follows that accounting for such inter-linguistically congruent items in lexical tests provides a more in-depth understanding of the multilingual lexicon.

Since Dijkstra et al. (2010) argues that the cognate facilitation effect serves as evidence
for a combined multilingual lexicon, the robust correlation between the cognate scores in the two languages also indicate that such congruent words are potentially integrated into the same system. Whether they are activated in a parallel fashion in both languages or just in the TL cannot be extrapolated from the results. For this can be task dependent, one can only tentatively assume that if the cognate is mastered in one language, it is also recognised in the other language in an automatic fashion. Students’ answers on the questionnaire confirm this as most of them found the instruction on cognates useful and reported that they do not consciously search for meanings in their L2. The above-mentioned pertains to a between-subject analysis, so as to further discuss the factors that influence lexical dexterity and their impact on vocabulary testing per se, the by-item findings are focalised next.

The proportion of cognates approximately match in both studies, comprising 56% of cognates in S2 and 53% in S1. The results obtained in S1 indicate a sharp contrast between cognate recognition in the L2 and L3. In English, significantly more cognates were recognised than non-cognates, however, in Romanian no such disparity was observable. While in Chapter 4 it has been argued that the cognate facilitation effect observed could be entrenched in the fact that English cognates might be easier for participants than the Romanian items, especially if the UWL is contrasted with the Romanian 6k level, Chapter 5 brings forth convincing evidence against this contradistinction. Although cognates gained a higher item facility index on both the ENVST and ROVST, the difference did not yield statistical significance. Additionally, the small effect sizes in both languages were around .10, which means that cognates in the two languages performed similarly. At least two issues emerge from this finding. First, the selection of the type of cognates can have a major impact on how the results are interpreted. Second, despite more students answer cognates correctly than non-cognates, the cognate facilitation effect is not obvious regarding the overall scores. From a language testing perspective however, it is also important that individually the items perform in accord with the expectations.
In relation to S1, the item analysis revealed regarding the English test that out of the six items that elicited a perfect score, five were cognates. 15 further items elicited an item facility index above .90 out of which only four items were non-cognates. Moreover, 66% of the items had a facility index above .50, and out of these, 37 were cognates. This is considerable as it confirms that 77% of the cognates were answered by more than half of the participants correctly. Compared to this, 20 non-cognates were below the .50 mark and 22 were above, indicating that non-cognates were more equally spread on the difficulty continuum.

Similarly, in the case of S2, the English test indicates that 66% of cognate items were answered by more than half of the students, whereas non-cognate items are equally spread above and below the .50 mark. Moreover, out of ten items that were answered by over 90% of testees, seven are cognates. On the RomVLT, ten items were answered correctly by more than 90% of participants and six of these are cognates. Additionally, 65% of cognates were answered by more than 50% of the participants. On the ROVST, out of nine items that were correctly answered at least by 90%, only one is non-cognate, and 62% of cognates have been answered correctly by at least half of the participants. Statistically speaking, a larger sample size could have potentially reflected these tendencies, thus teasing apart the discrepancy between cross-linguistically congruent and incongruent items and bolstering the conclusions.

Nevertheless, the above interpretation of scores from an item-analysis perspective may serve as convincing evidence for the cognate facilitation effect and emphasises the importance of accounting for cross-linguistic similarities in lexical research. Notwithstanding the fact that only recognition tests were operationalised here, it can be speculated that this will be prevalent in production as well. If cognate items are generally answered by more participants and thus, they are easier to be recognised, this might also mean that such items will demonstrate increased lexical availability. Hence, they will be
recalled or produced in association or word elicitation tasks, such as the Lex30, at a greater rate. Considering the fact that interlingual homographs are of lower frequency in English, this would enable learners to attain higher scores on similar tests, thus overestimating their VS. This issue has not been empirically addressed yet in studies employing the Lex30 (Fitzpatrick, personal communication).

The two studies also revealed that overall the items might be easier for students, especially in the case of cognates. In the case of the Hungarian test, even though the initial idea was to compare RT measures only between the L1 and the two additional languages, items should have been selected perhaps from only mid-frequency ranges and up to 20k to obtain a more even distribution and enable the comparison of lexical proficiency as well. Since Hungarian is an agglutinative language, even this might not have resolved the issue though. Conversely, the lower the word frequency in Hungarian, the more scientific or academic words would occur, and therefore be sampled, which in the case of Romance language speakers and university students could still elicit high scores and risk a ceiling effect.

At the other end of the scale, by examining the items that elicited answers from only a minority of students, admittedly, there are two problematic characteristics to notice. The analysis revealed that the cognate items that obtained a very low score had either a misleading distractor, in the case of S1, or a non-defining sentence with a potential prime in S2. Whereas such non-cognate items in S1 elicited no responses and all of them pertained to the lowest frequency band, in S2, only two items were non-cognates and these did not display any discrepancies.

This, on the one hand, reiterates that cognates in lexical tests should be regarded as special items, not just for the interlingual similarities they share in students’ background languages, but due to potential distractors that could become misleading given the language constellations that are examined. On the other hand, some of the cognate items were simply
not part of the participants’ lexicon and, despite cognates attained a higher item facility ratio in comparison to non-cognates, it is evident that not all cognates were recognised across the board to the same extent, which is demonstrated by the individual differences.

Similar findings were reported by Nagy et al. (1993) and they also suggest that cognates that were not recognised might not be known in students’ background languages either. Although the studies herein preclude drawing conclusions on whether cognates were not recognised just in the test language, or not known in the other language either, it seems reasonable to assume, based on the low frequency of misfitting items, that they did not form part of students’ lexica. On the contrary, the fact that not all cognates have been recognised by all students indicates that it cannot be taken at face value that cognates are equally recognised or known across languages. This idea has been investigated through RQ1 (6.1).

6.5 The interpretation of test scores

The tests’ limitation with regard to inferring far-reaching conclusions about actual vocabulary knowledge have been addressed in various publications (e.g. Milton, 2009; Milton and Fitzpatrick, 2014; Nation, 2001; Read, 2000; Schmitt, 2010). The most recent and comprehensive perhaps is Kremmel and Schmitt (2016) which investigated two different facets of vocabulary tests. The extent to which learners can actually use or comprehend the correctly answered items in production or reading is the first one. The second one is concerned with the depth of knowledge and whether the results from such formats allow inferring far-reaching conclusions of learners’ conjugational or derivational knowledge, awareness of collocational or contextual deployment of the items that they answered correctly on the test. Their findings suggest that neither the VLT (multiple-matching), nor the VST (multiple-choice) format is suitable for predicting meaning recall knowledge or the items’ employability in other modalities. There are two issues indicating that these findings have to be treated as tentative though.

First, the participants of their main study were German NSs. Therefore, it is safe to
assume that based on the format of the items, in the case these are cognates, learners can recognise the corresponding meaning. However, it comes naturally that knowing the meaning of the words cannot be compared to derivational/conjugational or collocational knowledge. For instance, the example item they provide is *beagle*, which besides being a cognate in German and English, is general-knowledge specific. Thus, students’ general knowledge about this word might not extend beyond the fact that it is a canine. Consequently, while the meaning is recognised in the matching formats as the definition contains ‘dog’, it is not certain whether the definition ‘*a small dog with short legs and long ears, used in hunting*’ (2016: 381) serves as enough information in the form recall format, regardless whether it is given in German or English, for those non-dog-loving participants. Arguably, all test items should be constructed so that they are culturally, linguistically, and general-knowledge independent, thus if an item favours cynophiles or others, its validity can be compromised.

Based on the above, it may be argued that the recall formats are problematic by their nature and could test something other than FL knowledge. Everyone is familiar with instances when one can describe a word perfectly, knows the word itself, but cannot produce it even in the L1. Unfortunately, since cognates have not been accounted for in Schmitt and Kremmel (2016), one cannot assume their distribution on the different formats employed in the study, and therefore, it is impossible to extrapolate the magnitude of their effect. Furthermore, Laufer and Goldstein (2004) present compelling evidence towards the fact that as language knowledge develops incrementally, the different types of word knowledge (or strength: form recall or recognition, and meaning recall and recognition in Schmitt’s terms [2010]) that map onto different test formats can be construed as a hierarchy. However, the strong correlation between the different formats indicates that while the four test formats produce significantly different scores, there is also a tentative relationship between them (Laufer and Goldstein, 2004).
Second, in comparison to the VLT and VST, Kremmel and Schmitt (2016) featured nine items in the four different test formats, 36 items in total. While as a small-case exploration of how different items function in a test, this would be informative if more item-characteristics were principally controlled for, for a quantitative study investigating over- or underestimation on the different formats, the number of items might be misrepresentative, especially if cognates are not accounted for. The studies described in the present thesis pointed towards the issue that partially due to the high number of cognates on the VLT, the tests were unable to establish language dominance. Moreover, it is emergent, in conjunction with previous studies, that cognates are not just specific items in learners’ lexicons but can impact testing and, therefore, should be accounted for accordingly.

The criticism offered by Kremmel and Schmitt (2016), and others in chapters 2 and 3, regarding the overestimation of the matching format do not undermine the results presented in this thesis, however. One reason for this is offered by Laufer and Goldstein (2004): if language knowledge is chiefly form-meaning driven and there is a clear difference in active (meaning) and passive (form) recognition, testing these independently can reveal information about the other types of word knowledge as well (recall) due to the associative relationship between them.

The other reason for this is that the underlying construct that has been investigated is the form-meaning mapping skill from a multilingual perspective. The interest was in comparing lexical recognition in the L2 and L3, and examining how this is impacted by cognates as a bridge between the two languages. Therefore, despite the fact that such test formats might overestimate examinees’ vocabulary knowledge due to guessing or elimination, supposedly, the extent of this pertains to the results obtained in both languages. Moreover, the results have not been deployed to infer about item employability in different modalities, what concerned me predominantly is how Romance languages can and do influence learners’ test scores and lexical assessments in general due to lexical similarities.
Such test formats have been speculated to overestimate the vocabulary knowledge of those who speak a Romance language (e.g. Nation, 1983; Schmitt, Schmitt and Clapham, 2001; Milton, 2009; Laufer and Ravenhorst-Kalovski, 2010), albeit the exact nature of this overestimation is as yet unaddressed. This issue is further discussed in the following.

Despite the fact that the two studies accounted for cognates in the tests and investigated the effect to which such items might influence the interpretation of lexical knowledge in a meaningful way, the exact nature of this overestimation has not been directly addressed thus far. The methodology sections highlighted that since the exact distribution of cognates between English and Romanian, or any Romance language for that matter, on different frequency levels have not been comprehensively investigated, the Romanian tests have followed rigorously the proportion of cognates on the original English tests. Whether this proportion represents the exact distribution of cognates in ‘real-language’ is considered an important factor from a testing perspective. Meara (1992), Eyckmans (2004), and Elgort (2013) propose that in order not to compromise the tests’ validity, the sampling should include a realistic number of cognates and these should be controlled. To allow for a quantified control of this overestimation, the number of cognates on the tested frequency bands should closely represent the proportion of cognates in English.

In order to determine the proportion and distribution of cognates, Romanian, Spanish, and Hungarian cognates within the most frequent 10k English words (BNC/COCA; Nation, 2006) have been explored in a follow-up study. Since this was done a posteriori of designing and conducting the present studies, this information could not serve as a source for test compilation. Nevertheless, it is considered that this could inform future studies in the area and facilitate their evaluation in retrospect.

Figure 6.1 below represents the preliminary findings of this research. This demonstrates that based on the applied criteria, among the most frequent 1k words of English, there are approximately 280 cognates, this increases to around 500 and slightly
above on the succeeding bands, except on the 3k band, which consists of 720 potential cognate words, and, thus, comprises the highest number of potential cognates. Regarding false cognates, the number is stable around 30 words per band. This provides a rough estimate of the proportion of cognates that should potentially be sampled from each of these levels.

Figure 6.1 Distribution of words on the BNC/COCA frequency list by Romanian cognate status

By rough estimate it is meant that cognateness is an abstract measure, as the previous chapters have highlighted. In particular, polysemous words can be regarded as cognates in some contexts, but not in others. Furthermore, cognateness not being a dichotomous measure is supported by that cognates can fully overlap in form and meaning, or just partially, especially if conjugational and derivational affixes are regarded in differing languages. Therefore, whether an English word is a cognate depends on the context and also the criteria that one applies. Nevertheless, the above indicates that for instance if ten words are sampled from each frequency band, then approximately three words should be cognates in the case
of band one, five to six in the case of all other bands, with the exception of band three where this number should increase to seven. Moreover, the entire test should contain at least 3% of false cognates evenly distributed on each frequency level.

The question, then, becomes how do the VLT and VST abide by this proportion? The VLT, for example, on the 2k band comprises ten cognates out of 18 words. Simply put, if ten items were selected from the 1,000 words of the 2k band, five should be cognates, with 18 items, roughly nine should be cognates. Thus, the VLT on this band has one extra cognate. On the 3k band, only ten items are cognates instead of 13. Similarly, on the 5k band the test should have nine cognates, instead of only five. Despite the fact that overall, there is only 1% difference between the total number of cognates on the frequency list and the number of cognates on the VLT, due to the frequency effect, the individual differences on the frequency levels could seriously impact the scores.

As this brief analysis demonstrates, due to the lower number of cognates on the test, it can be argued that the VLT in fact is likely to underestimate Romance language speakers’ (my research shows that the same percentages are roughly true for Spanish as well, with individual item variations) receptive vocabulary knowledge. From a multilingual perspective though, if the fact that knowing a word’s meaning from a background language does not mean that learners are familiar with the items’ all derivational and inflectional forms, might not be able to use or recognise it in different contexts, and also, might not be aware of its collocations or the appropriate register (e.g. Nation, 2013) is taken into account, it can be corroborated that actually the tests overestimate Romance language speakers’ overall proficiency. The overestimation in the case of English and Romanian, based on the overall percentage of cognates could probably reach 50%, given the right context or definition on the test. This further supports previously made claims: cognates represent special items for lexical tests and these should be controlled in the case typologically similar languages, where the proportion of affordances they provide could be much more substantial. In turn, this
would allow for a more in-depth understanding of the lexicon, especially in the case of
multilinguals. However, the issue of cognateness is also affected by the complexification
caused by the frequency of words. Frequency and cognacy of words, due to their reciprocal
impact and mutual contribution to the lexicon, therefore, seems to demand control in not just
the TL, but the background languages as well.

The high proportion of cognates between Latinate languages and the English language
serves as further evidence for cognates having a facilitative effect not just during the
acquisition process, but in testing as well. Moll et al. (1992) state that cross-linguistic
congruency can provide learners a common lexical heritage enabling them to capitalise on
cognates as a source of knowledge. Regarding academic words, Cummins (1994) intimates
that knowledge of concepts from the background languages potentially advance these
concept-laden words’ acquisition. The above-described proportion of cognates in a Romance
language and English, the distribution of cognates on the vocabulary tests, and the results
seem to support this. In particular, S1 demonstrated that the UWL provided participants an
advantage compared to the same amount of cognates on the RomVLT 6k band that were not
academic words. Moreover, as has been demonstrated, cognates triggered a higher item
facility ratio overall, which suggests that such words are easier for learners who are familiar
with a Latinate language, even if it is not their L1 and their proficiency varies. This lends
some support in the direction that the study has a number of didactic implications, discussed
in the upcoming sections.

6.6 The relationship between speed of lexical access and vocabulary size

The proportion of cognates in the TL does not merely impact lexical knowledge but,
as S2 reveals, there is a robust relationship between lexical access and VS. Additionally, S2
also corroborated a facilitation effect of cognates in terms of speed of response. This is in
line with several studies that have demonstrated a close relationship between lexical access
and the degrees of similarity between certain words (e.g. Elgort, 2011; Lemhöfer et al.,
Chapter 6: Discussion

2004). RT studies have also shown that the more proficient a learner is in a language, the faster they respond to items in a variety of tasks, thus RT can also indicate language proficiency or dominancy (Lemhöfer et al., 2004). Furthermore, cognates have been proven to be recognised faster and more accurately than non-cognates, especially in the non-native languages. In spite of this, the employment of a temporal measure in vocabulary studies has been scarce. S2 has purposefully been designed in such a way that it adds not just to the aforementioned aspects of proficiency in the L2 and L3, but also contributes to the currently scant knowledge in the area of vocabulary studies in relation to RT.

The underlying assumption behind employing a temporal measure was that speed of lexical access, in conjunction with vocabulary scores allows the investigation of other important aspects of vocabulary knowledge (Meara, 1992; Miralpeix and Meara, 2014). It was also considered that as an objective quantifier, the RTs with a form-meaning recognition task can be replicated in similar research (Tanabe, 2016). Furthermore, by assessing participants in three different languages, it was hypothesised that proficiency should be reflected by the speed of lexical access, which in turn becomes a contributive component to our understanding of the interrelation of multiple VSs. The logical entailment of this is that in an integrated ML, the lexical links between L2 word forms are naturally weaker and reduced in number, compared to L1 links (Meara, 1992). Therefore, in a task that explores lexical knowledge and then, consequently, the number of these links, measuring the speed of access is well-justified, insofar as, this is surmised to resemble the strength of such links. In other words, if VS attempts to describe the lexicon in terms of number of units that one knows in a given language, RT should be able to reflect on how quickly one can access these units, and presumably, the more units are stored and the stronger the links are, the faster the responses should be.

These hypotheses have been partially borne out by the evidence in Chapter 5. In accord with this, the RT measure confirmed that the higher the students’ vocabulary score is in a
language, the faster they respond. This confirms findings in previous studies (e.g. Aparicio and Lavaur, 2014), despite the methodological differences, and corroborates that RTs decrease as a function of increasing proficiency across languages. Naturally, students obtained the lowest RTs and answered most accurately in the L1. Respectively, students gained significantly higher scores in the L2, which was reflected in their mean RTs and lower disparity figures as well in comparison to English, in which they attained the lowest scores and the highest RT.

Consequently, in spite of the differences in the test format (compared to yes/no tests in e.g. Harrington, 2006), and the technical operationalisation (using a mouse, instead of buttons in e.g. Tanabe, 2016) the RT measures obtained apparently align with proficiency estimates and confirm the expectations. This echoes previous studies suggesting that RT is faster in the more dominant languages. The other two issues that were investigated through the temporal measure, was the size and frequency effect in relation to lexical access.

In full accord with results from earlier studies (e.g. Harrington, 2006; Laufer and Nation, 2001), the findings in S2 also extrapolated that students with an increased lexical proficiency provide faster answers than testees with a lower proficiency. This was true, however, only in English and can be explained as follows.

It seems rational that the lower the proficiency in a language and the greater the disparity between the individual students, the more this manifests in their RTs. Therefore, as students are more proficient in Romanian and the disparity between the two groups’ proficiency scores is smaller (still significant though), the less probable it is to find a difference in the speed of responses on a multiple-choice format, where a non-defining sentence and four options are also provided as these have the potential to seriously increase the RT. If the item is more abstract or difficult, students, according to the anticipations, require more time to make a decision. This suggests that perhaps the format of the test overshadows some important aspects that a binary lexical decision task would still allow to
be teased apart, which can be partially considered as a limitation of the current methodology. However, it also obtains that currently no RT measure provides a true reflection of real time processing (e.g. Tanabe, 2016) and thus, employing the multiple-choice format that produces relative RT appears to be valid.

The other concomitant issue that has been investigated is the frequency effect in relation to RT. Harrington (2006), Laufer and Nation (2001), Tanabe, (2016), inter alia, proclaimed that speed of lexical access increases inversely proportionally to word frequency. It can be corroborated from S2 that this stands in relation to both English and Romanian. As a general tendency, in comparison to the first 2k most frequent words, items on lower frequency levels were answered slower. This indicates that accuracy, as a reflection of the process of decision making, and RT demonstrate well to what extent proficiency and word frequency influence the outcome.

Given the multilingual perspective in this thesis and the RT measure, it has also been of particular interest whether cognates are recognised faster than non-cognates. Surprisingly, a difference was only found in the L1 and L2, but not in the L3. The mean RTs in English suggest, however, that even if the difference in means is very small, cognates are recognised faster and more consistently than non-cognates. This is in line with previous findings (e.g. Brenders et al., 2011; Elgort, 2011; Van Hell and Dijkstra 2002) and a number of other studies that arrived to the conclusion that cognates have a facilitative effect as they are not just responded to more quickly, but in turn they also facilitate reading, comprehension, and fluency as well (e.g. Otwinowska, 2016; Szubko-Sitarek, 2015).

Despite this, the fact that cognates were recognised faster in the L1 than non-cognates is diverging from findings in Van Hell and Dijkstra (2002) and Brenders et al. (2011), which conclude that the cognate facilitation effect is not a determinant factor in the L1. The reason for this difference probably rests in the fact that the proportion of cognates in the HUVST was balanced according to the number of English-Romanian cognates in the VST, which
most probably inflates the proportion of cognates in Hungarian. Conversely, the majority of cognates in Hungarian are common English and Romanian words, very representative of academic words as well: e.g. \textit{concentrate}, \textit{paragraph}, \textit{forum}, \textit{portrait}, \textit{nationalism} etc. Thus, it becomes partially clear why such words would have a strong link within participants’ ML.

Finally, linking the faster responses to cognates back to the ML, Lemhöfer et al. (2004) intimates that cognates would not show a facilitative advantage in terms of RT if language access was selective. This supports further the aforementioned findings that point to an integrated lexicon, where cognates are a combinatorial node in an intertwined system.

\textbf{6.7 The effect of frequency and other word characteristics}

While the effect of frequency largely holds true in both studies and in both languages, not just in terms of VS, but also RT, there also seems to be compelling evidence that word occurrence is not the sole factor that underlies form-meaning recognition, and by extension, lexical acquisition. English vocabulary, as a source of language knowledge, appears to be not just influenced, but to some extent, also determined by prior lexical knowledge. The effect of cognates seems to facilitate word recognition, which, in turn, makes vocabulary testing problematic as the postulations in the present chapter have implied from different perspectives. S1 exemplifies this best.

Cognates were not only answered more accurately, but the inclusion of the UWL in the test further conflated the scores. Moreover, in the case of both studies, the item facility indices reveal that cognate items elicit higher number of responses, and thus, can be considered easier. These findings echo previous assertions in the area of bilingualism and L3 acquisition, albeit often not properly addressed in lexical studies.

A limited number of studies explored other word properties that can influence VS measurements. Milton and Daller (2007), for example, reported only the frequency effect as being significant and concluded, with some reservations, that this is the best explanatory variable of vocabulary scores compared to cognateness. Corroborations in Willis and Ohashi
(2012) and Reynolds et al. (2015) point to a noncommittal stance. Whilst the former one found that the cognateness of items explains more variance in the scores than frequency of occurrence, the latter one found that polysemy and frequency contributed approximately equally to the model, and cognacy was non-significant. These are in contrast with the current findings as the mixed-effect models in S2 yielded no effects for POS or word length operationalised as the number of syllables. With the exception of Milton and Daller (2007), where French intermediate learners have been tested on the yes/no format, the two studies reported above and S2 utilised the VST. Subsequently, the discrepancies in the findings point towards alternative speculations than test formats.

One obvious point to consider is learners’ background languages. Willis and Ohashi (2012) tested Japanese learners and the proportion of cognates was fixed at 29%, whereas Reynolds (2015) tested Taiwanese English majors and the number of cognates was reported as around 2%. In the present case, for Romanians the number of cognate affordances was over 50%, however, cognates emerged only in the case of VLT and the low-proficiency groups in S2 as significant predictors. It follows from this that the proportion of cognates and learners’ background languages probably influence whether cognates yield a facilitative effect or not. Additionally, this might be further influenced by the frequency of cognates in the L1 and L2, as proven by Laufer and McLean (2016).

The other point to consider is learners’ proficiency. Willis and Ohashi (2012) report VS scores for their participants within the range of 2k-5k, while in S2 the minimum score was 3,800 with an average of 8,200. This makes it amenable to theorisation that students with a lower proficiency in Willis and Ohashi could benefit more from the facilitative effect of cognates, and hence, it emerged as a significant predictor of the model, compared to S2. The fact that the Low group’s score was significantly influenced by cognate scores lends additional support to this. Overall, it appears that while the effect of frequency can be reconciled across the board, with certain background languages and proficiency levels,
cognates can potentially exacerbate the conceptualisation of word knowledge and the subtleties of score interpretation.

Importantly, it can be stipulated herein that lexical scores tend to demonstrate propensity towards the frequency effect. All test results clearly indicate a disparity between high- and mid-frequency vocabulary scores, however, it is entirely conceivable that the traditional convention of dividing the most frequent words into 1k bands deserves reconsideration and this should be espoused for further explorations in successive studies.

The data collected reveal that on the VLT, the difference between 2k and 3k words was non-significant. While on the mid-frequency ranges a decreasing tendency was certainly observable, the contrast is only moderately accentuated. Consequently, perhaps future test compilations could take into account the proficiency of learners and alter the number of words within each band according to that. High-frequency bands (up to 3k) could be amalgamated into one band and subsequent bands could be broken into 500-word bands to enable (by also increasing the number of tested items) a more detailed learner profile. Along similar deliberations, Kremmel (2016) suggested that high-frequency bands should be divided into 500-item bands, while mid-frequency bands could retain their current size (1k). Admittedly, this would only serve beginner and intermediate learners as it would allow researchers to explore their lexical profile in a greater depth at high-frequency ranges. Whichever the case, reconsidering the current number of items in a frequency band could give way to an array of new research opportunities and with properly accounting for cognate items, and perhaps controlling for the frequency effect in all languages in the DLC, test scores could be adequately calibrated and better interpreted.
Chapter 7: Conclusions

Part of my aspiration at the onset of this doctoral project was to attempt to engineer vocabulary tests that can uncover particular attributes of the multilingual lexicon, and the RomVLT and ROVST emerged as a result of that. However, now, following two empirical studies and subsequent discussions of the findings, it appears that working backwards and reconceptualising some of the conventions that define lexical tests will be perhaps more revealing for future research. One of the primary reasons for this is that an explanation of the size of the whole ML, or a particular TL, which relies merely on the frequency of word occurrence, it seems, is clearly untenable in researching multilinguality.

The known words in multiple languages and the similarity between these, give the impression that an in-depth understanding of the target lexicon may be impossible, or at least highly misleading for researchers and practitioners alike, if multilinguals’ ML was not embraced as a dynamic and composite entity (e.g. De Angelis, 2007; Otwinowska, 2016). This is due to the fact that each of the linguistic subsystems, or previously learnt languages, influence each other and are highly intertwined at various levels, a conception that is under-developed in vocabulary testing literature.

Measuring VS in two typologically closer languages and accounting for cross-linguistic similarities offered solutions to this problem. Receptive vocabulary knowledge has been previously conceptualised as a core contributory factor to LA, and within this, other component language skills, such as reading or writing. A secondary impetus for this research was therefore to develop a systematic approach to the analysis of the multilingual lexicon, expose problematic aspects of vocabulary knowledge, and propose probabilistic explications of the key issues that govern lexical testing.

In order to achieve this, a learner-centred framework was adopted, which, retrospectively, enabled the collation of information on the participants that could be compared against certain language-related issues, thus complementing the extrapolations on
the lexicon and learners’ perception of their spoken languages to a finer level of granularity. Specific internal and external validity and reliability criteria were employed to scrutinise the findings in both an a priori and a posteriori fashion.

The remainder sections provide an account of the procedures and methodological decisions involved in the study, and describe its findings, contributions, and implications. No research is without limitations, however. These form, then, the second part of this conclusion. Finally, key take-away messages are offered with some concluding remarks.

7.1 Reflections on the contributions and implications

Despite the complex and multi-faceted nature of a multilingual lexicon, at this point, it is reasonable to state that on a theoretical level this thesis problematised and explored two broad facets that contribute to the field of vocabulary research and, consequently, has obvious implications. The first is multilingual vocabulary knowledge and the role of cognates as a special bridge or catalyst between languages, which in turn relates to the pedagogical implications of this thesis. The second concerns vocabulary testing and the issue of cognates from a test design and score interpretation perspective, which indicates the methodological implications.

7.1.1 Pedagogical contributions and implications

The most direct link between the current research and teaching or learning vocabulary is the experimental paradigm, i.e. the EI on cognates. Both studies indicate, and confirm Molnar (2010) in that, given the age and proficiency of the groups under study, the cognate awareness-raising exercise had no effect on either the overall VS or the accuracy of recognising cognates of those in the SGs. However, this is far from implying that cognates should not be part of the (Romanian) language teaching curriculum either explicitly or implicitly. On the contrary, these learners should be seen as the ‘end result’ of primary and secondary school teaching. Therefore, understanding how linguistically competent university students’ lexicon comprises their L2 and L3, and what the relationship is between
these have clear value for developing more fine-tuned and nuanced teaching methods that focus on comprehension.

Since participants of this research are undergraduates of English and will reportedly leave academia as translators, interpreters, language teachers and so on, probably, it is safe to assume that they represent individuals who can function reliably in English, and therefore, should embody a tentative ultimate attainment for learners of lower proficiency. Teaching a FL in two or three hours a week is a challenging task however, especially for large groups with varying abilities and cognitive skills. Consequently, exploring methods that facilitate LL has values from a didactic and cost-benefit perspective.

Of significance is the finding that there is a robust relationship between Romanian and English lexical knowledge, and this is further accentuated in the case of written cognate recognition. Furthermore, the fact that cognate knowledge in one language is a strong predictor of overall vocabulary knowledge in the other lend way to speculation that cognates form part of the same conceptual system. In other words, since cognates are purportedly stored in the lexicon in an integrated fashion, this could mean that learners can capitalise on such affordances without the explicit teaching of cognates and, in turn, such words can facilitate reading and comprehension. Based on previous studies and the Parasitic Model, this is expected for advanced, adult learners and disproves the first hypothesis within this cultural context.

Regarding lower-level learners, however, Morán-Molina (2010), for example, advocates the deployment of lexical similarities in the form of cognate-induced reading tasks. According to him, increasing the number of cognates in a text gives learners confidence in their language skills, provides a sense of familiarity with the text and, broadly speaking, allows them to focus more (i.e. reallocate valuable resources and time) on other important aspects of LL, especially high-frequency non-cognates. Consequently, the lack of a significant effect of the EI herein is certainly no entailment for disregarding the benefits of
cognate awareness-raising tasks at low-proficiency levels.

Other studies involving younger or less proficient learners (e.g. García, 1991; Jessner, 1999; Otwinowska, 2016; Proctor and Mo, 2009; Ringbom, 2007) also intimate that teaching cognates has the potential to increase comprehension and metalinguistic awareness, which facilitates LL. Therefore, it is entirely conceivable that the emergent results only pertain to the highly proficient participants tested or to the lexical tests employed that are not sensitive enough to detect the benefits of EI on cognates. This is further substantiated by the fact that on the questionnaire over 90% of the participants claimed that they rely on their L1 when searching for a suitable English word, instead of Romanian, and that most students in the SGs, with some reservations regarding their L2 knowledge, found the instruction on cognates helpful. Thus, it becomes amenable to theorisation that cognates, as a catalyst between certain languages, can contribute to language education and similar research could inform multilingual approaches to teaching.

Central to such research projects is the fact that they are limited in most cases to a small number of language aspects or target groups. In a broader context though, if the number of Romance language speakers as an L1 is taken into account and the omnipresence of English in almost all areas of younger generations’ life, it becomes probable that the present findings have far wider-reaching implications. Considering that the topic of the current research is the acquisition of English as an L3 and the number of speakers of Latinate languages as an L2 or Ln is immense, it seems reasonable to imply that cognates may serve as lexical resources for a large group of learners, which further justifies the Focus on Multilingualism initiative propagated by Cenoz and colleagues (e.g. Cenoz, 2013; Gorter and Cenoz, 2016). The fact that in the present case students are heritage speakers of Hungarian and learn Romanian (to a varied extent, depending on the area they grow up in, family background, education, predisposition towards the state language, perception of its importance etc.) from an earlier age, in a greater number of school hours and, presumably,
are exposed to it to a larger extent than English, means that due to the large number of
cognates, their L2 is a potential lexical source they can rely on from the initial stages of their
LL journey. Molnar (2008), inter alia, serves as proof for this as Hungarian-Romanian
bilinguals outperformed their Hungarian monolingual peers in an English lexical task, which
contributes further support to the benefits of cognates.

By way of instantiation, Latinate cognate words, such as possible (RO: posibil; SP:
possible); interest (RO: interes, interesa; SP: interes, interesar); area (RO: areal; SP: área);
present (RO: prezent, prezenta; SP: presente, presentar); subject (RO: subiect; SP: sujeto);
experience (RO: experienţă; SP: experiencia); member (RO: membru; SP: miembro), which
are all within the 500 most frequent words of English and, therefore, represent essential
words that provide a large coverage in general discourse, will present a challenge to
Hungarian, Chinese or speakers of other languages that are neither of Latinate nor Germanic
origin. However, as the examples demonstrate (but examples from widely-spoken Romance
languages, such as French, Italian, Portuguese could also be added), the orthographic
similarity between the forms of these words enables learners with a background knowledge
of these languages recognising these lexemes instantaneously, perhaps without much
cognitive effort. Although it is entirely tenable that such words will not yet form part of
young learners’ conceptual lexicon, in the case of cognitively developed (i.e. more adult
learners) at the onset of their LL trajectory, such words will provide a reliable lexical crutch
when reading or even in language production. Generalising possibly even further, as studies
such as Bartolotti and Marian (2017) instigate the effect of wordlikeness as beneficial to LA,
it appears that the common fallacy of the negative impact of false friends dissipates concerns
about the degree of meaning overlap or deceptive nature of interlingual homographs.

It also follows from this that English NSs can also rely on such lexical affordances
when learning a Romance L2. However, perhaps more importantly, this could aid teachers
in guiding their students, in Cobb’s words, ‘where, on the open seas of a second lexicon,
particular learners could most usefully cast their nets’ (2002: 296). His research found a clear advantage for cognates in the case of French monolinguals – students were more inclined to check Anglo-Saxon words in the dictionary than Latinate items, and knowledge of non-cognate items barely contributed to reading comprehension in comparison to cognates. This serves as further evidence for cognates contributing to not just lexical knowledge on tests, but might also aid reading. If this is conjoined with assertions by Bartolotti and Marian (2017), it lends way to conjecture that purely formal similarities of words between different languages can probably facilitate the formulation of associative or semantic links, thus contributing to and accelerating the acquisition and long-term retention of such lexis. While it is acknowledged that the exact way cognate knowledge transfers to spoken comprehension and production is much less clear, this thesis focused exclusively on written meaning recognition.

Unravelling the rate at which learners acquire new lexis and the route of vocabulary development has been at the heart of SLA research for years. Tapping into this particular aspect is largely masked by a number of factors, such as quantity and quality of exposure to the TL, learning needs, age and motivation of learners, and as the present line of enquiry indicates, learners’ background languages in the case of typologically less distant languages. Milton (2009) and Read (2000), amongst others, reveal that although on a group level learners’ profile follows and can be best described by the frequency of occurrence of words, on an individual level they demonstrate differences and learn some unusually infrequent words prior to acquiring the most basic vocabulary. It appears that one explanation for this could pertain to LL being economically driven - learners rely on language similarities and attempt to find patterns in the interlanguage to reduce the learning effort (e.g. Hall et al., 2009). Therefore, cognate knowledge can be tentatively conceptualised as one of the potential sources for the so far unexplained proportion of this coincidental knowledge in learners’ lexical profile (e.g. Milton, 2009).
Chapter 7: Conclusions

The orthographic similarities in the above examples and other cognate items also lend way to speculation that learners, either intuitively or purposefully, rely on such similarities and, consequently, mark such items as known on a yes/no test or find their meaning without much effort in a multiple-choice or matching test, provided that the item is known in their Ln.

Strongly attached to this argument is the oft-mentioned observation that Romance cognates in English are of lower frequency than in a Romance language. Therefore, for Romance learners of English the meaning of a cognate may be easier to grasp (e.g. Lubliner and Hiebert, 2011; Nagy et al., 1993). This encapsulates a straightforward implication for pedagogy, and indicates that there might be clear value in explicitly focusing on the development of such metalinguistic skills in the case of less advanced and young learners. This is because their Romance background is a potential fund of knowledge of even more abstract concepts, thus reducing the learning burden.

It also follows that the evidence lined up herein in favour of the Interdependence Hypothesis rationalises that developing vocabulary knowledge benefits not just the L2, but the L3 as well. The broader the students’ lexicon is in Romanian, the broader it seems to be in English. Taking into account that neither in the L2 nor L3 were all cognates recognised, indicates that there is room for improvement though. The methodology employed in the present thesis enabled to meaningfully compare lexical knowledge in the two languages, and one clear implication is that the size of students’ vocabulary is not only related in the L2 and L3, but lexical knowledge can also be interpreted as a function of academic achievement. Thus, increasing Hungarian NSs’ lexicon in Romanian may be transferable knowledge to other languages, such as English or Spanish, and in turn, this could also facilitate the development of other skills (e.g. reading) and, consequently, benefit academic success.

The results also imply that speed of lexical access is in a reciprocal relationship with language proficiency as the negative correlations reveal. Since Laufer and Nation (2001)
link accessibility directly to language fluency, this suggests that the more proficient learners are in a language, all other things being equal, the faster they recognise items or access the form-meaning link. In other words, it appears that a large lexicon facilitates the search between form and meaning, and this can directly influence reading. Generally, both studies in the current thesis confirm that lexical proficiency could be a core element of language development on multiple levels and if academic success is interpreted as the amalgamation of other linguistic, metalinguistic, and communicative competencies, relying on cognates can potentially enhance LL and reduce the cognitive effort of acquiring new lexis or even deploying cognates in the TL.

Another, perhaps thus far slightly deemphasised, contribution also deserves attention. SLA research is largely occupied with the acquisition of English and this is further exacerbated by the fact that apart from bilingual vocabulary tests, monolingual tests in non-English languages hardly exist (see Daller and Ongun, 2017 for a Turkish X-Lex; Milton, 2009 for French; Milton, Alexiou, and Mattheoudakis, 2014 for Greek). As Chapter 3 has elucidated, this is also a lacuna in the Romanian context. The RomVLT and the ROVST appears to address this paucity in a seemingly valid and reliable way, which, in turn, also implicates that the RWL developed for the present thesis can be utilised for developing tools and materials for L2 learners of Romanian. The tests can be deployed at any age and proficiency and used to monitor learners’ progress or divide them according to proficiency. This contribution is of relevance as normally Hungarian NSs are compared to Romanian NSs on language tests that employ the same marking criteria, which places L2 speakers of Romanian at a disadvantage.

Inquiring into how many words are known by L1 speakers of different backgrounds would allow establishing the necessary thresholds for comprehension and text coverage at certain levels. These are directions that would benefit not just Romanian L2 learners and teachers, but policy and material writers as well. The figures proposed in the present thesis
clearly follow and deliberately build on the accepted standards in English at our current knowledge. However, it is not impossible that these thresholds for high-, mid-, and low-frequency vocabulary have to be tailored to other languages accordingly. These studies serve as a starting point for such explorations and discussions, nonetheless, there is clearly more work to be done in this area.

Another implication that is of pedagogical relevance was revealed by the questionnaire in S1 and confirmed in S2. The triangulation between lexical scores in the L2 and L3, and the self-rating questions corroborate a strong association between the size of students’ lexica and their confidence and proficiency ratings in both formal and informal settings. Although the interpretation of confidence and proficiency on similarly short, self-rated questionnaires is fuzzy at best, as not all personality traits can be controlled for, it is certainly encouraging to see that a larger vocabulary also results in students feeling more confident and proficient in both languages. Notwithstanding the fact that such correlations do not corroborate a causal relationship, taking into account other possible and latent factors that similar studies cannot control, the probability of these relationships is valuable for future research.

The questionnaires also highlight that despite having grown up in Romania, students feel more proficient and confident in using English than Romanian. Perhaps, as they are students of English, self-awareness is biased by their aspirations. Nevertheless, this indicates that starting from an early age, Romanian educational institutions should reconsider the curricula and implement changes that foster vocabulary development and use in realistic settings with the hope of increasing learners’ metalinguistic and communicative competencies. In a heavily literature- and grammar-oriented education system, this might prove to be a challenging task though.

The questionnaires also revealed that the majority of students link their L3 to Hungarian, instead of Romanian. Furthermore, when reading, they either translate the English word into Hungarian or try to guess from the context, and no use of Romanian is
indicated. This elucidates that at least consciously, the L2-L3 link is very poor, which is further rationalised by the fact that 50% of students consider Romanian useful. Results like this might be confounded though by students’ predisposition towards Romanian or their perceived (psycho)typological distance between Romanian and English.

Although the current research showed some evidence towards an automatic search for form-meaning connections of cognates, it would be far too simplistic to conclude that each and every cognate that is known in the L2 or L3 will be intuitively recognised and, if this strategy fails, then students rely on their L1. This entailment is largely tentative and further substantiations are called for in the case of beginner and lower intermediate students to support the present findings. One clear implication herein is that students definitely find the EI helpful and the prevalence of the allegedly strong L1-L3 link requires further explorations. From this research it is impossible to conclude whether similar cognate awareness-raising exercises aid the retention or potentially encourage the use of cognates in alternative language modalities but, based on previous findings, the explicit time that is spent on increasing less proficient learners’ cross-linguistic awareness and vocabulary development will presumably not be in vain.

7.1.2 Implications and contributions to testing

The second cluster of implications is concerned with the issue of cognates from a testing point of view. An advantage of the employed methodology, which allowed focusing on a specific aspect of the ML, is that it yields precise results regarding the impact of other salient variables. In particular, by exploring the relationship between the L2 and L3 interlanguages, it became possible to account for cross-linguistic similarities pertaining to these languages on a lexical level and examine to what extent cognates impact vocabulary scores.

Both studies converge in the inference that interlingual homographs sharing a meaning in the L2 and L3 affect the item facility indices. In essence, cognates appear to be easier
items for learners and while this can be considered as a facilitative effect for LA, such items are liable to becoming problematic for testing. The ratiocination behind this is that cognates seem to interfere with score interpretation, pose a potential threat to the tests’ content validity, baffle cross-cultural comparisons of language learners and, subsequently constrain, if not obliterate, the possibility of standardisation of similar assessment tools.

Moreover, if a test was deployed to measure students’ progress in a course, it would be convoluted to interpret whether cognate items are known because of effective teaching and encountering the item frequently enough for retention, or because it was already part of their vocabulary. Alternatively, other wordlikeness metrics can also contribute to the ease of retention of congruent words, such as neighbourhood size or orthotactic probability. In similar cases, maybe even one encounter of the English word would enable students to make the link between the new target form and its meaning, whereas incongruent items would certainly need more iterations in accordance with other characteristics that influence word difficulty. Previous studies have provided sustenance to this and claim that cognates are recognised, learned, and translated faster than incongruous lexis (e.g. Otwinowska, 2016; Szubko-Sitarek, 2015). It is also conceivable that controlling for such lexical similarities in research that focuses on vocabulary uptake or explicit lexical teaching could illuminate the rate of passive vocabulary acquisition and its governing factors. Therefore, conceptualising an eloquent way to account for cognate knowledge in lexical tests seems to be a useful start. Hindrance to a solution in this direction derives from multiple aspects however.

Perhaps the most pervasive of these is operationalising cognates as a counting unit. Due to the fact that cognates naturally occur between certain languages, completely eliminating them from lexical tests appears to be counterproductive and contraindicated (e.g. Eyckmans, 2004; Meara, 1992). In order to maintain the results’ validity, Elgort (2013) and Meara (1992), inter alia, construe that cognates should be represented in a realistic proportion in tests. While section 6.5 provides a preliminary indication of the proportion and
distribution of cognates in English and Romanian, it remains elusive how such items should be scored.

As the aforementioned postulated, scoring is problematic even from a frequency point of view. Although in the case of high-frequency items it is relatively safe to assume that if one of the sampled words is known, then other items from the same frequency band will be potentially (but only potentially) known as well. However, with mid-frequency or less quotidian lexis this is increasingly problematic. Conversely, if one cognate is known, can we assume that other cognates from the same frequency band will be known? Probably, an affirmative answer to this is more tenable than in the case of non-cognates. Nevertheless, the results confirm that regarding cognate and non-cognate items equally on tests seems to implicate an oversimplification in multinguality as it masks underlying factors that pertain to the dimensions and development of vocabulary knowledge. Reflectively, these problems also contribute to the difficulties in test-score calibration.

Concern with the scoring of individual items on lexical tests also introduces a macroscale source of dilemma, which is the exact estimation of VS. Scores can potentially underestimate the size of the lexicon if cognates are eliminated or underrepresented. Antithetically, overestimation can occur if inter-linguistically congruent items are featured predominantly on a lexical test or in instances when the number of cognates is representative in the stratified sampling, however, far-reaching conclusions are drawn about the items’ employability in alternative contexts (e.g. knowledge of additional meanings, sentential deployment etc.).

Such implications in the case of homogenous groups that share the same or similar language background (DLC) are perhaps negligible. However, for cross-cultural examinations, such as admission tests or group allocations by proficiency in EAP or pre-university language courses, the consequences can be detrimental. One empirical instantiation of this is found in Helms-Park et al. (2009), who compared Romanian and
Vietnamese learners and found that cognates facilitated word recognition for the Romanian group, despite the fact that the two groups scored equally on non-cognate items. On the one hand, this suggests that in terms of receptive skills, the difference between the two groups might be reasonable. In relation to productive skills, on the other hand, the Romanian group’s knowledge may be overestimated though.

Concurrently, Laufer and McLean (2016) also support that the percentage of cognates in a lexical test can induce a difference in the scores and provide an advantage to learners of a more basic proficiency. The findings herein harmonise with these conclusions as cognates increased the likelihood of answering test-items correctly and lower-level learners could capitalise on such affordances. Despite many studies acknowledge that cognates should follow a realistic proportion in lexical tests and call for controlling this overestimation, rare are the studies that attempt to systematically investigate the effect of cognates or quantify their distribution in a principled way. Section 6.5 explores the proportion and distribution of Romanian-English cognates and false cognates on the BNC/COCA list and corroborates that, depending on the context, degrees of lexical and semantic congruence, Romanian learners’ VS might be overestimated by half of their actual score. This is because the list comprises around 50% of potential Romanian cognates, compared to typologically more distant languages, such as Hungarian, Vietnamese, or Arabic. While this might give the impression of being an extreme estimation, it has to be regarded as a probabilistic overestimation that is context, meaning, and task dependent. Let us look at a couple of examples.

Romanian, Spanish, German, and even Hungarian students recognise with ease the word bank (RO: bancă; SP: banco; GE: die Bank; HU: bank) on a lexical decision task or even attach the most common meaning to it (financial establishment). However, in a reading comprehension or productive task in which it is used with the meaning of land alongside a river, speakers of the above languages might have difficulties, unless they are familiar with
this particular meaning in English already. Similarly, Romanian and Spanish speakers potentially have no problem with recognising or retrieving the word *rest* (RO: rest; SP: resto) with the meaning of *remainder*. However, in the sense of *relaxation* or *pause*, if the English meaning is not part of their (conceptual) lexicon, the item might become deceptive. Even so to the extent that, if a non-defining sentence is provided and testees realise that a verb is expected (e.g. *I need to rest*) they might translate it as or choose the distractor *deduct* or *take away* based on the Spanish verb *restar*, in which case the item becomes an obvious false cognate (for Spanish speakers). Not so much so for a German speaker, as *der Rest* and *die Rast* or *rasten* carry both semantic alternatives (*remainder* and *pause* respectively). The cognate *fine* as an adjective might be a straightforward hit for Romanian and Spanish speakers (RO: fin; SP: fino), however, as a noun or verb in the sense of *financial penalty*, it might confuse learners, which points to the difficulties of interpreting yes/no tests or psycholinguistic lexical decision tasks. Since some views implicate that the human brain works on the least effort principle (e.g. Hall et al., 2009; Zipf, 1949) and, therefore, humans attach additional meanings to quotidian lexis, thus also increasing the polysemous nature of language(s), the above exemplification could be extended to a large number of cognates in numerous languages.

This implication elucidates why cognates are hard to define, quantify, and generate inherent problems for lexical assessments. The above examples highlighted the fact that English lexemes can have multiple meanings, albeit this is also the case in other languages, which further complicates controlling cognates in a meaningful way. This is premised in the principle that languages develop separately and the frequency of word occurrence is not ubiquitous across languages. For instance, Spanish and Romanian quotidian lexis might represent infrequent words in English. Additionally, the number of different meanings attached to such words can also vary across languages. It is also commonly agreed that not even NSs are familiar with all the meanings of a word or oversee the subtleties of differences
of synonyms for example. Consequently, it remains the case that research into multilinguality should systematically account for lexical similarities, which can strengthen the validity of lexical tests.

From these implications, it appears conspicuous that the engineering of a more nuanced, interlanguage-sensitive, and valid VS test that can stand the probe of standardisation will lie in the direction of resolving the issue of scoring items based on their cognate status and frequency in a more empirically substantiated way. This ostensibly simple task is complicated further by the extent of cognate recognition skills that students employ during lexical access and this might only be reliably captured by treating cognate-related and cognate independent skills and knowledge deliberately and severally.

Subsequently, if cognate recognition is reportedly a separate skill (Nation, 2013) and this is conjoined with the emergent discussions, it can also be concluded that bi- or multilingual learners, due to their plurilingual competence, are likely to be learners in their own right and should be treated differently to monolinguals acquiring an L2.

In sum, learners’ languages are apparently intertwined in the ML and, as such, these subsystems mutually influence each other. For this influence to be captured comprehensively and permit the researcher to decide where the actual knowledge derives from out of the multiple possible sources, it seems reasonable to implicate that multilinguals should be regarded as distinct subjects with specific skill-sets.

7.1.3 Raising the potential of a paradigm shift

Finally, it is plausible, both on a methodological and theoretical level, that the implications pertaining to this thesis warrant for a fundamental paradigm shift in vocabulary studies towards a holistic view of the ML. Within this holistic view it is regarded that the interconnection between the known languages necessitates an in-depth exploration of the intricacies of the ‘whole’ as a complex and dynamic entity. This, in turn, can guide the apprehension of the component parts. Therefore, it is proposed that empirical investigations
being directed towards the lexicon of multilinguals might encapsulate a viable panacea for uncovering and expediting how vocabulary is retained, recognised, and recalled in a specified TL, and how this aids the reconceptualisation of the size(s) of the lexicon. Ultimately, this would open up the enticing possibility of generating normative data for VS extrapolation in different language constellations in both the receptive and productive dimensions.

On a similar note, this would also concur with the unprecedented interest in multilingualism and L3 acquisition on not just the level of current language policies and research endeavours, but on a pedagogical level as well. Equally, a paradigm shift in this direction would entail the discovery of the relative importance of measuring the total vocabulary size and the total conceptual vocabulary of learners (see Core et al., 2013; Pearson, Fernandez and Oller, 1993).

These two notions found prevalence in bilingual child LA studies in order to account for the conception that in the case of bilinguals the meaning of a word (concept), might be known in the L1 or the L2, or both. Thus, in order to understand how many words a learner knows it is not their monolingual (NS) counterparts they need to be compared to, but their vocabularies have to be taken into account in conjunction.

Under this notion, then, bilinguals, or by extension multilinguals, are not at disadvantage as at the conceptual level, they know the same amount of words as their monolingual peers, and this is called the total conceptual vocabulary. If monolinguals and multilinguals’ total vocabulary size is taken into account, however, multilinguals are likely to know even more words than monolinguals if lexis is counted in all their languages severally. In other words, the total VS consists of all the words known by the individual and the conceptual VS comprises all the concepts in the ML, but only once. Thus, surprise, surpriză (RO), sorpresa (SP) and ajándék (HU) would be counted as one lexeme (concept) in the conceptual lexicon and four separate entries in the total vocabulary.
Chapter 7: Conclusions

The methodology employed in the current thesis is thought to be straightforwardly adaptable for a similar line of enquiry and could be employed to identify critical features of LA and testing in classrooms that encourage multilingualism and translanguaging to foster communication and nurture language and skill development. The issue of cognates is definitely one of these critical features and the applied method seems to be informative. Nevertheless, limitations form a vital part of research, and it is important to detect these as they can contribute to and, if necessary, veer future research directions. The following penultimate section will address these.

7.2 Reflections on the limitations and future directions

Practical, methodological, and some a posteriori limitations have been explicitly acknowledged and described in the preceding segments, therefore, this section focuses on some general limitations. Amongst the most obvious and, perhaps, inevitable ones that pertain to the methodology these can be mentioned: sample size of the population, choice of vocabulary tests and languages, range of proficiency levels and age, and the dimensions of lexical knowledge investigated.

7.2.1 Limitations relating to the population

The participants of the present thesis only represent a small cohort of language learners with a Romance language background and pre-defined proficiency level. This limitation could have been overcome in two different ways. Ideally, S2 should have addressed either learners whose L2 is a different Romance language (e.g. Spanish or Italian) and study English as an L3. This would have been problematic, on the one hand, due to potential interference from the L1, which in the case of Hungarian speakers was reduced to a minimum and, more importantly, systematically controlled for. On the other hand, the availability and comparability of monolingual Spanish or Italian form-meaning mapping lexical tests (such as the VLT or VST) and my knowledge of these languages would have unavoidably and adversely impacted the time constraints allowed for this project.
Alternatively, the number of participants and the range of proficiency levels covered, as a limitation, could have been overcome by administering tests to all students from Partium Christian University or by approaching other institutions. Both of these solutions have been attempted. However, testing the entire student population was not feasible due to facility issues (a single computer laboratory). The State University in Oradea has been approached as well, but due to the differences in teaching periods and availability of the gatekeeper, the data collection turned out to be unachievable. These two solutions would have clearly enabled the inclusion of higher proficiency Romanian speakers (compared to their English) and beginner to high intermediate English learners (not English majors), thus increasing the range of explored proficiency levels. The inclusion of younger learners has also been considered for a possible third study, but due to external and beyond-control constraints, such as the lack of time and resources, this has become infeasible.

Increasing the sample size to above 100 would have not just contributed more statistical power to the studies, but would have also allowed the fitting of a Rasch model or a three parameter logistic (3PL) model from Item Response Theory to the data. Either of these models, or in combination, would have facilitated a more in-depth analysis of the relationship between person abilities, test items and their difficulty, discriminative power, and guessability. These models bear the appellation of ‘modern’ test theory as they attempt to rectify the attenuation of limited interpretative value of reliability metrics offered under Classical Test Theory. Nevertheless, the reliability measures (including Cronbach’s alpha, and inter-item and total-item correlations) and item facility indices proved to be valuable in evaluating the results. The latter revealed that cognates are not just easier items that influence scoring, as prominent, these also provide an advantage to low-proficiency students and, conversely, can become misleading or deceptive items for Romanian speakers.

Probably, a larger number of participants could have also led to a more meaningful analysis between cross-linguistically congruent and incongruent items and could have
revealed the true difference in means, which is perhaps masked here by the very high proficiency of participants. In other words, young or less proficient learners comparatively might have been able to capitalise more on the L2-L3 lexical similarities, thus leading to a statistically significant difference between cognates and non-cognates. Nonetheless, as established in the preceding sections, S1 and S2 unveiled quintessential features of the multilingual lexicon in the case of students who can be regarded as ‘the end result’ of primary and secondary education. These features, in turn, have the benefit of informing future directions of exploring and assessing multilinguals’ VS.

Additionally, examining the relationship between the L2 and L3 in the case of different proficiency levels would have enabled extrapolating whether the strong link between lexical knowledge obtains to less proficient learners as well. This would have further substantiated the arguments laid out herein for an integrated lexicon. From a testing point of view, this would have also lent additional support to the overall conclusion that test items should be cross-culturally considered and scrutinised to ameliorate the validity of the test’s results, and the recommendation offered in relation to the reconsideration of score calibration in terms of frequency and cognate status of items. The virtue of this is that it could result in more robust, interlanguage-sensitive tests and, perhaps, would be an essential step towards standardisation.

7.2.2 Limitations relating to the dimensions of vocabulary knowledge

Throughout this thesis, it has been emphasised that lexical knowledge is complex and multifaceted, which is even further complicated if a multilingual dimension is added. Despite most studies reviewed herein, including the two empirical studies conducted, analyse receptive lexical knowledge through frequency-based vocabulary breadth tests that operationalise the form-meaning mapping knowledge, this aspect cannot necessarily be regarded as a limitation per se. The reasons for this are as follows.

First, as Chapters 2 and 3 highlight, lexical knowledge in general is a strong indicator
of language proficiency overall, the different receptive and productive modalities (writing, reading etc.), grammatical knowledge, and as the current findings confirm, it has a strong explanatory power of academic success, cognate recognition, and speed of lexical access.

Second, previous research has shown (e.g. Gyllstad, 2007; Laufer and Goldstein, 2004) that lexical recognition is a significant predictor of other types of vocabulary knowledge (e.g. form or meaning recall) or progression on the CEFR levels (Milton, 2010; 2013), therefore, by exploring one component part, it becomes possible to extrapolate more in-depth on the ML or language proficiency as a whole. Since LA is gradual in its nature and within this, receptive knowledge develops at a faster rate than productive knowledge, and passive VS tends to be larger than active (e.g. Milton, 2009; Schmitt, 2014), understanding the intricacies of meaning recognition in the case of multilinguals makes viable the possibility of exploring and exposing other characteristics that govern or could expedite lexical development or recognition.

Third, the aim of language processing and production is largely meaning driven (e.g. Meara, 2010; Milton, 2009; Schmitt, 2010). Language usage, simply put, is chiefly communicating meaning in some way or form, and understanding what is communicated to us. In this view, uncovering certain characteristics of the form-meaning mapping knowledge, as a clear and consequential notion for language users, can probably be seen as the most essential trait of LA. The tests deployed to investigate this trait are not without fault however.

7.2.3 Limitations relating to the tests employed

The problem with multiple-choice or matching tests is that no realistic situation presents learners with clearly defined options to select the correct meaning from (Schmitt, 2010). While speaking or writing, language users have to rely on words that are readily available to them in order to convey meaning, and often these words belong unexclusively to the TL. In reading or listening, meanings can be deducted from the available conceptual knowledge of the given word or contextual clues. This way, presenting testees with a range
of possible options is further complicated by the lack of context, correct understanding of
the provided responses and distractors, elimination, or cognate recognition skills. In spite of
these problems, many argue that tests such as the VLT and VST are the most proximate tools
in the direction of standardised tests (e.g. Schmitt, 2010; Webb and Sasao, 2013).

Furthermore, exploring the ability of attaching the correct meaning to an item, or
passive form recognition on a lexical decision task, as it were, says little about the extent to
which learners can comprehend or deploy such words in different language modalities.
Equally, productive tests are largely confined by the type of task or the context researchers
employ to elicit words. Thus, it appears that instead of posing a limitation for the current
empirical studies, the format of the tests used, based on the wealth of questions they address
in research and available validation studies to the present date, these should be rather seen
as useful tools for exploring the VSs of the multilingual.

7.2.4 Limitations relating to item characteristics

Two other, perhaps more profound limitations have to be mentioned regarding the
conceptualisation and operationalisation of lexis from a frequency and cognate perspective.
These two facets are strongly interlinked. The first, frequency, is important because, as S1
and S2 revealed, the frequency distribution and proportion of cognates on the different bands
was determined by the design of the original English tests. Although a preliminary analysis
rectified this limitation in section 6.5 to some extent, and demonstrated that while both tests
should undergo some minor adjustments in terms of cognate distribution, the overall impact
of this was only minimally detrimental and should be further examined.

Alternatively, the frequency of word occurrence in the background languages as an
effect on accuracy, lexical access, or scoring should also be explored as a potential
explanatory variable for measuring VS. Cognate frequency in Romanian might prove
revealing in explaining recognition in English. Future studies examining the relationship
between a Romance language and English have the possibility of building on such
limitations and addressing these accordingly. The rationale for this is that in order to ensure
the tests’ validity is not, or just negligibly, compromised, these should closely match the
percentage of cognates in ‘real language’ and also consider the calibration of cognate and
mid- to low-frequency items’ representativeness and scoring, thus corroborating or falsifying
the proposed extent of overestimation in the present thesis.

The second facet concerns the choice of individual test items in the two empirical
studies. The similarity of interlingual homographs, with or without a meaning overlap
between the known languages, should be rather seen on a continuum and as such, a
dichotomous classification of cognate and non-cognate items potentially masks some
underlying traits of the lexicon as a multimodal and versatile resource for vocabulary
knowledge and language production.

A more specific classification of cognates from an orthographic, phonologic, and
semantic point of view, therefore, could shed light in future studies on the retrieval and
accessibility of similar linguistic affordances in lexical tests. However, as S2 shows,
alternative word-characteristics such as POS or word length do not affect word recognition
on the employed test formats neither in terms of accuracy nor RT. Thus, it becomes
conceivable that classifying cognates according to their formal similarity, polysemy, or other
characteristics would still not yield a significant effect in the case of highly proficient
learners. The logical entailment for this reasoning is that the time and cognitive process of
the form-meaning attachment is mainly governed by students having to read the non-
defining example sentence and the four multiple word definitions. This tentatively intimates
that in the case of multilinguals, employing different test formats, such as a mixture of
translation or retrieval tasks with more refined incarnations of the VST could lead to a better
understanding of the interconnections in the ML.

Furthermore, cognates in all languages under study represent distinct items on the tests.
Therefore, it becomes problematic to extrapolate that knowledge of a congruent item in the
L1, L2, or L3 results in recognising the similarity or meaning intuitively in the other languages. Based on the item facility indices, the speed of lexical access data, the lack of effect of the EI, and the strong correlation between the cognate scores tend to suggest though that if the item is known in one language, it is likely to be recognised in the other one as well. Ensuing studies in the area could revisit this aspect by sampling cross-linguistically identical or similar items in all tested languages. If the majority of these items are recognised in all languages, this would imply that items are activated automatically and confirm the current suggestions that regardless of the language used to elicit the meaning of such words, if the item is known, it will be responded to correctly in any language (non-selective access view). Antithetically, diverging findings and discrepancies in the case of certain items could indicate that the lexeme was either guessed in the language where it was correctly answered or some other aspects (e.g. degree of similarity, polysemy, difference of item frequency in the tested languages etc.) inhibit the cognate facilitation effect, insofar as, they represent different lexemes in the background languages (selective activation).

Lexical transferability, that is, the probability of an item (cognate or non-cognate) being transferred to the TL (Kellerman, 1983), can be constrained or purposefully inhibited by a number of additional factors: (psycho)typological distance between the languages, linguistic factors, cumulative LL experience and knowledge, metalinguistic or paralinguistic factors, psychological factors (cognitive, attentional, and developmental), learning environment, frequency of use etc. (Jarvis and Pavlenko, 2007). Attempts to unmask these factors in conjunction with the size of the lexicon and transferability provide the possibility of a more comprehensive understanding of the lexicon and attainment, retention, inhibition, or even fossilisation of lexis.

Closely related to the issue of cognates is the lack of representativeness of deceptive or false cognates in the tests in both studies. The ratiocination for this is premised in (1) the realisation that the negative, inhibitory, and perplexing view of false friends in research and
classroom practices is based on a common fallacy (Morán-Molina, 2010; Pons-Ridler, 1984) and (2) that the original VLT comprised no false cognates. In conjunction with Morán-Molina (2010), in section 6.5 it has been indicated that false cognates are present in Romance languages and English in a disproportionately small number (3%) compared to roughly 50% of (potential) cognates, depending on context. Thus, it seemed appropriate to aim to explore the facilitative effect of cognates in both studies. Conversely, even if false cognates in the tests would approach the approximate proportion in a frequency list, the limited number of such items would still prohibit drawing far-reaching conclusions on a quantitative basis. Follow-up studies should, however, inquire into the extent to which false cognates affect word and cognate recognition, whether students are aware of the distinct meanings in their L2 and L3, or the presence and proportion of such lexis in productive language modalities.

### 7.2.5 Limitations relating to research design

The last two limitations afforded here are also design-related aspects. One is the operationalisation of the temporal measure in S2 that allowed the investigation of lexical access in addition to VS in a multilingual context. The other one is the absence of a post-hoc test either in written or semi-structured interview formats, which could facilitate a more detailed exploration of multilingual lexical knowledge.

The RT measured in relation to the multiple-choice format can be considered problematic because, contrary to lexical decision tasks with a binary format, these are potentially influenced by not just word characteristics (e.g. length, animacy, etc.), but by the length of the non-defining sentence and the four options to choose from (and therefore, by extension, reading speed and comprehensibility of the item in the given language [i.e. the existence of a translation equivalent or similar concept in the Ln] can also contribute to this). In the case of less quotidian lexis (i.e. less frequent or non-cognate), arguably, some students can ponder or hesitate more to submit the final answer, while others will be more confident or care less to make the same decision. This way, students’ language proficiency and
personality both contribute to the actual time they take to submit an answer. Consequently, these relative RTs are not directly comparable to reaction times elicited by lexical decision tasks, albeit still proved to be meaningful.

Another limitation perhaps that generally relates to the multiple-matching and multiple-choice testing formats is that, from a cognitive processing perspective, little is known about the underlying processes that characterise such test-taking operations and their cognitive validity. In other words, it would be enlightening to understand which part of a test item learners concentrate on most, whether they focus on the non-contextual sentence while searching for meaning or they rather consider the options several times, maybe even eliminate until only two options remain or simply just guess, and how these issues all relate to cognateness and word frequency inter alia. There is convincing evidence from eye-tracking studies that successful test-takers tend to focus more on areas of key interest, that is, where the correct answer lies (e.g. Bax, 2013). Predictably, the eye-tracking method in conjunction with the temporal measure would enable invaluable empirical demonstrations with theoretical nuance in uncovering the cognitive processes of lexical test-taking using similar formats, a direction that is worth considering in future studies.

Nevertheless, based on Laufer and Nation (2001) and Tanabe (2016), the data obtained here revealed comparatively some scarcely addressed aspects of lexical access and confirmed some of the posed hypotheses. In particular, it appears that VS in multiple languages, as a function of lexical access and proficiency, can feed directly into the currently scant knowledge of the size of the multilingual lexicon and speed of meaning recognition on multiple-choice tasks. Future studies have the opportunity of using the present findings as a comparison in terms of speed of meaning recognition in different background languages and frequency levels within those languages, maybe expanding to other measures, such as the coefficient of variance to understand more about automaticity and lexical fluency.

Finally, there is convincing evidence, including the empirical findings uncovered in
this thesis, regarding cognates constituting a problem in recognition as not all learners automatically see them as cross-linguistic or translation equivalents and, therefore, their presence in different tasks can often go unnoticed or tend to be avoided (e.g. Dressler et al., 2011; Nagy et al., 1993; Otwinowska, 2016; Schmitt, 1997). A study exploring vocabulary learning strategies by Schmitt (1997) and the results on the questionnaire in sections 4.8.4 and 5.8.5 indicate that checking for cognates in the background languages is not something that learners consciously or deliberately do. One of the limitations of both studies in this thesis is the lack of a post-test or semi-structured post hoc interviews that could have allowed an inquiry into this issue. For instance, a post-test using three-way translations (L1 to/from L3; L1 to/from L2; L2 to/from L3) would have revealed to what extent students see interlingual neighbours as translation equivalents in these languages. Similarly, semi-structured interviews, as recommended by Schmitt (2010), are an excellent method for validating test results and exploring in more detail students’ lexical knowledge and processing of the tested items. Although the experimental paradigm, on the one hand, and the questionnaire on the other hand, had the objective of addressing these phenomena, in retrospect neither of them fully subscribe to the issue of cognates directly.

While the EI on cognates unravelled that proficient adult learners do not benefit from such interventions, or at least this is undetectable by meaning-recognition tests, students on the questionnaire stipulated that they mainly rely on their L1, instead of the typologically closer language to English, and found the instruction on cognates helpful. Although these findings have implications for future research, admittedly, either of the abovementioned post-hoc exercises would have provided more groundwork for the conclusions offered in this thesis. Future studies can build on and benefit from such limitations and systematically investigate these methodological issues depending on the context and the RQs they bid to address.

Overall, contrary to these limitations, it becomes evident that the future direction of
vocabulary research offers greater ecological validity for classroom assessments if the background languages, especially in the case of (psycho)typologically less distant languages, are accounted for. Inquiring into the size of the multilingual ML through the adoptable approach employed and explored in this doctoral thesis definitely answered some questions, but in the same time also managed to raise new, perhaps innovative, and exciting, avenues for future empirical endeavours that can reach beyond the monolingual speakers’ L2 VS and study the lexicon as a storage and multimodal resource for LA and production. This, in symbiosis, would benefit learners and educational stakeholders, thus enhancing the rate and possibly improving the experience of LL.

7.3 Final conclusions

This thesis offers a systematic approach to comparing vocabulary sizes in typologically close languages on a lexical level, such as Romanian as a member of the Romance language family and English, which is principled, robust, and encapsulates a wide range of analytic possibilities that are expandable and adaptable. This furthers the current paucity in how lexical knowledge in more than one interlanguage is incorporated and can be accounted for in the ML. It is clear, however, even from this learner-centred multilingual framework that the strengths and limitations of such approach to research will be largely determined and justified by the extent to which one can depart from and challenge traditionalised conceptions and assumptions.

The doctoral research reported in this thesis is the culmination of a motivated journey that started in 2013 and strived to connect two thus far disparate research fields: vocabulary studies and multilingualism. En route, it tentatively confirmed, contributed to and, hopefully, initiated some innovative and valuable ideas and discussions that can be summarised as follows:

a. Multilinguals, especially with a Romance language background, should probably be considered differently to monolingual learners acquiring an L2
b. Vocabulary knowledge and the examination of this, by necessity, is only approximate and deliberately reduced in complexity, however, through exploring one dimension (i.e. written form-meaning recognition in the dominant languages), it seems, important issues can be uncovered.

c. Lexical knowledge in the L2 and L3, on different levels, may be interlinked and mutually influential.

d. Frequency of word occurrence is repeatedly reconfirmed as an invaluable descriptor of lexical acquisition and a prominent test-design criteria, however, apparently, it is not the sole explanatory variable of these.

e. Cognates, as a crutch or potential affordances for learners and a bridge between certain languages, can transcend these limits and emerge as problematic items in testing.

f. Frequency and cognateness in conjunction, and if properly quantified, can potentially increase test accuracy, thus making an essential step towards standardisation and, perhaps theorisation.

g. Highly proficient learners seem to recognise cognates intrinsically, however, this cannot serve as evidence for lower level or younger learners not benefitting from EI on inter-linguistically congruent items.

h. Learners with a broad vocabulary recognise more cognates than those with smaller vocabularies, although, it appears that lower level learners can benefit from their cognate recognition skills, thus reducing the gap compared to their more proficient peers.

i. Technological advancements can be effortlessly employed in classrooms and research practices to measure lexical accessibility, and, in turn, even as a relative measure, RTs emerge as informative descriptors of vocabulary knowledge in multiple languages: there is an association between speed of recognition and proficiency in different languages; within each FL, lexically more able students also tend to answer faster; highly frequent items and cognates (in the L1 and L2) are likely to be answered faster.

j. Depending on the questions, apart from VS and RT estimations, a learner-centred framework should involve a questionnaire exploring participant-related factors, which, in exchange, may project relevant information on otherwise inexplicable research aspects.

This foray and the emergent issues regarding multilinguality and cognates will
expectedly inform, add value to, and form the basis of my own and others’ research and pedagogical agendas in a similar vein. Hopefully, readers are convinced that the present journey has been a productive and useful one in problematising and discovering some of the quintessential attributes of vocabulary acquisition and measuring vocabulary size.
References


Jessner, U. (2014). On Multilingual Awareness or Why the Multilingual Learner is a Specific Language Learner. In M. Pawlak and L. Aronin (Eds.), *Essential Topics in Applied*


Appendices
The English Test

This is a vocabulary test. Please choose the right word to go with each meaning. Write the number of that word next to its meaning.

Here is an example:

1 business
2 clock  6 part of a house
3 horse  3 animal with four legs
4 pencil  4 something used for writing
5 shoe
6 wall

If you have no idea about the meaning of a word, do not guess. But if you think you might know the meaning, then you should try to find the answer.

2000 Level

1. original
2. private  ____ complete
3. royal  ____ first
4. slow  ____ not public
5. sorry
6. total

1. apply
2. melt  ____ choose by voting
3. jump  ____ become like water
4. manufacture  ____ make
5. elect
6. threaten

1. blame
2. hide  ____ keep away from sight
3. hit  ____ have a bad effect on something
4. invite  ____ ask
5. pour
6. spoil

1. accident
2. choice  ____ having a high opinion of yourself
3. fortune  ____ something you must pay
4. debt  ____ loud, deep sound
5. pride
6. roar
Appendix 2

Extract from the RomVLT

2000 Level

1. factură
2. companie
3. vechime
4. asistent
5. festival
6. amintire

1. clădire
2. intensitate
3. rinichi
4. supraviețuire
5. dialog
6. semnătură

1. soț
2. alergie
3. mulțime
4. orientare
5. sfat
6. valută

1. stabilit
2. paralel
3. descrescător
4. dulce
5. natal
6. util

1. repeta
2. împiedica
3. pregăti
4. angaja
5. estima
6. părăși

1. sosi
2. facilita
3. încetă
4. cuceră
5. acuza
6. interesa
Appendix 3 The Questionnaire (Study One)

The following questions will help us understand more of your language background. Please read the questions carefully and answer them honestly. There are no right or wrong answers.

1. At what age did you start learning each of these languages?
   a. Romanian ______ b. English _____ c. Other (please name) ____________

2. On a scale of 1 to 5, please rate (by circling the number) how confident you feel using Romanian in everyday situations (e.g. restaurant, shopping, chatting to friends).

   1 2 3 4 5

   Not confident Fairly confident Very confident

3. Please rate how confident you feel using Romanian in a formal setting (e.g. job interview, presentation, business meeting).

   1 2 3 4 5

   Not confident Fairly confident Very confident

4. Imagine you work at an international company and are responsible for emails, phone calls, meetings etc. Which option would allow you to carry out your duties more professionally? Think of professionalism or efficiency, not preference.

   a. I would definitely feel more professional using Romanian
   b. I would feel more professional using Romanian in most situations
   c. I would be equally professional in Romanian and English
   d. I would feel more professional using English in most situations
   e. I would definitely feel more professional using English

5. Imagine you moved to an English-speaking country. Please rate how confident would you feel using English in everyday situations (e.g. calling locals, ordering food, going on a date with a local).

   1 2 3 4 5

   Not confident Fairly confident Very confident
6. You are writing an essay in English. The right word/expression is just not coming to your mind. Which of the following describes you best? In most cases:
   a. I think of suitable words in Hungarian and translate them
   b. I think of suitable words in Romanian and see if they work in English or translate them

7. Thinking of your Romanian knowledge. How much would you say it helps you in learning or functioning (speaking/writing) in English?
   a. It definitely helps
   b. In many cases it comes useful
   c. Sometimes it helps, sometimes it doesn’t
   d. Doesn’t make a difference at all
   e. It can be confusing and misleading

8. If you received the handout on cognates before the study, please answer the question below. If not, please turn over 😊
   How much did the handout on cognates help in making a connection between English and Romanian? How much do you think it would help / would have helped if you received similar instructions more often when learning English? How important/useful you think learning English through Romanian instead of Hungarian would be? Please comment in a couple of sentences (approximately 100 words).

Your input in this study will shape our understanding of the links between Romanian and English, and what this means to learners in general. Thanks a lot for your time and participation in the study. 😊 Please turn over.
As you may already know, there are quite a few words in English that resemble their Romanian equivalents (or vice versa). These words are labelled under the general term of cognates. Empirical studies have shown that cognate recognition can help students expand their vocabulary, understand and analyse texts more efficiently, and overall, infer meaning and use the target language more proficiently.

Please consider the following examples and make note of any additional ones that you remember.

1. Orthographically identical cognates

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>control</td>
</tr>
<tr>
<td>agent</td>
<td>agent</td>
</tr>
<tr>
<td>social</td>
<td>social</td>
</tr>
<tr>
<td>test</td>
<td>test</td>
</tr>
<tr>
<td>exemplar</td>
<td>exemplar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>uniform</td>
<td>uniform</td>
</tr>
<tr>
<td>participant</td>
<td>participant</td>
</tr>
<tr>
<td>minor</td>
<td>minor</td>
</tr>
<tr>
<td>gastric</td>
<td>gastric</td>
</tr>
<tr>
<td>epic</td>
<td>epic</td>
</tr>
</tbody>
</table>

2. Identical stem/base word but different ending

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>interview</td>
<td>interviu</td>
</tr>
<tr>
<td>expire</td>
<td>expira</td>
</tr>
<tr>
<td>remark</td>
<td>remarca</td>
</tr>
<tr>
<td>register</td>
<td>registra</td>
</tr>
<tr>
<td>juvenile</td>
<td>juvenil</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>station</td>
<td>stație</td>
</tr>
<tr>
<td>pronounce</td>
<td>pronunța</td>
</tr>
<tr>
<td>vote</td>
<td>vota</td>
</tr>
<tr>
<td>affinity</td>
<td>afinitate</td>
</tr>
<tr>
<td>stabilization</td>
<td>stabilizare</td>
</tr>
</tbody>
</table>

3. Regular orthographic correspondences such as: -y / -ie, -ence / -ență, -able / -abil

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>disability</td>
<td>dizabilitate</td>
</tr>
<tr>
<td>eligibility</td>
<td>eligibilitate</td>
</tr>
<tr>
<td>decadence</td>
<td>decadență</td>
</tr>
<tr>
<td>preferable</td>
<td>preferabil</td>
</tr>
<tr>
<td>melancholy</td>
<td>melancolie</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>sequence</td>
<td>secevență</td>
</tr>
<tr>
<td>susceptible</td>
<td>susceptibil</td>
</tr>
<tr>
<td>independence</td>
<td>independență</td>
</tr>
<tr>
<td>convergence</td>
<td>convergență</td>
</tr>
<tr>
<td>misery</td>
<td>mizerie</td>
</tr>
</tbody>
</table>

4. Irregular orthographic relationships and orthographically or morphologically more distant words

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>complicated</td>
<td>complicat</td>
</tr>
<tr>
<td>honour</td>
<td>onor</td>
</tr>
<tr>
<td>chemistry</td>
<td>chimie</td>
</tr>
<tr>
<td>ethnography</td>
<td>etnografie</td>
</tr>
<tr>
<td>hereditary</td>
<td>ereditar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>engineering</td>
<td>inginerie</td>
</tr>
<tr>
<td>magic</td>
<td>magie</td>
</tr>
<tr>
<td>ballet</td>
<td>balet</td>
</tr>
<tr>
<td>included</td>
<td>inclus</td>
</tr>
<tr>
<td>monarchy</td>
<td>monarhie</td>
</tr>
</tbody>
</table>

Although we have to acknowledge that there are some false cognates (or false friends), such as eventual, library, preservative, sensible and so on, our knowledge of either English or Romanian can actually facilitate learning and speaking the other language. Please bear this in mind during the tests.
Appendix 5

Consent Form

All the information you provide will be anonymised and stored securely. The data will be analysed and reported in such a way that no individual can be identified and no one will know who participated in this study. Consent forms are only accessible to the researcher named above and will be used and stored in compliance with the Data Protection Act.

Please tick the boxes below to indicate you read and understand the statements, and sign at the bottom.

☐ I am willing to take part in this research, and I give my permission for the data collected to be used in an **anonymous** format in any written reports, presentations and inclusion in published papers relating to this study.

☐ I understand that my participation is voluntary and that I am free to withdraw within two weeks after today’s date.

☐ I understand that the results of this study will be made available to me upon my request in form of published article, conference paper or any other published or unpublished form.

Your name and year of study:

Signature:

Today’s Date:

**Thank you for your participation!**
## Item Table for the ENVST, ROVST, HUVST

<table>
<thead>
<tr>
<th>L</th>
<th>POS</th>
<th>English W</th>
<th>C</th>
<th>S</th>
<th>Rom W</th>
<th>S</th>
<th>C</th>
<th>Hun W</th>
<th>S</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>V</td>
<td>SEE</td>
<td>NC</td>
<td>1</td>
<td>AVEA</td>
<td>2</td>
<td>NC</td>
<td>TART</td>
<td>1</td>
<td>NC</td>
</tr>
<tr>
<td>A</td>
<td>POOR</td>
<td>NC</td>
<td>1</td>
<td>ÎNTREG</td>
<td>NC</td>
<td>VALÔ</td>
<td>2</td>
<td>NC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>STANDARD</td>
<td>RO - HU</td>
<td>2</td>
<td>PROZÁ</td>
<td>2</td>
<td>EN - HU</td>
<td>PÂRT</td>
<td>1</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>PERIOD</td>
<td>RO - HU</td>
<td>3</td>
<td>CERTIFICAT</td>
<td>4</td>
<td>EN</td>
<td>PROGRAM</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>SHOE</td>
<td>NC</td>
<td>1</td>
<td>CER</td>
<td>1</td>
<td>NC</td>
<td>SOR</td>
<td>1</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>N</td>
<td>STONE</td>
<td>NC</td>
<td>1</td>
<td>SEARĂ</td>
<td>2</td>
<td>NC</td>
<td>REMÉNY</td>
<td>2</td>
<td>NC</td>
</tr>
<tr>
<td>N</td>
<td>NIL</td>
<td>NC</td>
<td>1</td>
<td>CAL</td>
<td>1</td>
<td>NC</td>
<td>MÚLT</td>
<td>1</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>MICROPHONE</td>
<td>RO - HU</td>
<td>3</td>
<td>OCAZIE</td>
<td>3</td>
<td>EN</td>
<td>GAZDA</td>
<td>2</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>UPSET</td>
<td>RO</td>
<td>2</td>
<td>RECE</td>
<td>2</td>
<td>NC</td>
<td>SZEGÉNY</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>CIRCLE</td>
<td>RO</td>
<td>2</td>
<td>MEMORIE</td>
<td>4</td>
<td>EN - HU</td>
<td>FÓRUM</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N</td>
<td>SOLDIER</td>
<td>RO</td>
<td>2</td>
<td>IRONIE</td>
<td>4</td>
<td>EN - HU</td>
<td>PUSZTA</td>
<td>2</td>
<td>RO</td>
</tr>
<tr>
<td>N</td>
<td>DINOSAUR</td>
<td>RO - HU</td>
<td>3</td>
<td>BIOGRAFIE</td>
<td>5</td>
<td>EN</td>
<td>RÉGIÓ</td>
<td>3</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>ROVE</td>
<td>NC</td>
<td>1</td>
<td>ÎNVINGE</td>
<td>3</td>
<td>NC</td>
<td>JELLEMEZ</td>
<td>3</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>SCRUB</td>
<td>NC</td>
<td>1</td>
<td>IERTA</td>
<td>2</td>
<td>NC</td>
<td>TÁMAD</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>DASH</td>
<td>NC</td>
<td>1</td>
<td>IVI</td>
<td>1</td>
<td>NC</td>
<td>RAK</td>
<td>1</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>COMPOUND</td>
<td>RO</td>
<td>2</td>
<td>FARMACIE</td>
<td>4</td>
<td>EN</td>
<td>KOCSMA</td>
<td>2</td>
<td>RO</td>
</tr>
<tr>
<td>N</td>
<td>INPUT</td>
<td>RO - HU</td>
<td>2</td>
<td>ŞOFER</td>
<td>2</td>
<td>EN - HU</td>
<td>PRÓBA</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>TUMMY</td>
<td>NC</td>
<td>2</td>
<td>PODEA</td>
<td>2</td>
<td>NC</td>
<td>KÚSZÖB</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>LATTER</td>
<td>RO</td>
<td>2</td>
<td>GROZAV</td>
<td>2</td>
<td>NC</td>
<td>REMEK</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>CRAB</td>
<td>RO</td>
<td>1</td>
<td>ADULT</td>
<td>2</td>
<td>EN</td>
<td>GRÔF</td>
<td>2</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>N</td>
<td>DEFICIT</td>
<td>RO - HU</td>
<td>3</td>
<td>SUBJEC TIV</td>
<td>3</td>
<td>EN - HU</td>
<td>STADION</td>
<td>3</td>
<td>EN - RO</td>
</tr>
<tr>
<td>N</td>
<td>COMPOST</td>
<td>RO - HU</td>
<td>2</td>
<td>BRONZ</td>
<td>1</td>
<td>EN - HU</td>
<td>KONSZENZUS</td>
<td>3</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>PEE L</td>
<td>NC</td>
<td>1</td>
<td>RÂNÎ</td>
<td>2</td>
<td>NC</td>
<td>RÔGZÎT</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>HOUNT</td>
<td>NC</td>
<td>1</td>
<td>SCURGE</td>
<td>2</td>
<td>NC</td>
<td>ERED</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>FRACT URE</td>
<td>RO</td>
<td>2</td>
<td>TRIUMF</td>
<td>2</td>
<td>EN</td>
<td>VÂM</td>
<td>1</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>PREMIER</td>
<td>RO - HU</td>
<td>3</td>
<td>PLATFORMÂ</td>
<td>3</td>
<td>EN - HU</td>
<td>OBJEKTUM</td>
<td>3</td>
<td>EN - RO</td>
</tr>
<tr>
<td>N</td>
<td>THESIS</td>
<td>RO - HU</td>
<td>2</td>
<td>GLUCÖZÁ</td>
<td>3</td>
<td>EN - HU</td>
<td>PARAGRAFUS</td>
<td>4</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>VEER</td>
<td>RO</td>
<td>1</td>
<td>POPULA</td>
<td>3</td>
<td>EN</td>
<td>KONCENTRÁL</td>
<td>3</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>BUTLER</td>
<td>NC</td>
<td>2</td>
<td>POVARĂ</td>
<td>3</td>
<td>NC</td>
<td>ÜZEMANYAG</td>
<td>4</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>STRANGE</td>
<td>RO</td>
<td>2</td>
<td>FORȚA</td>
<td>2</td>
<td>EN</td>
<td>FORMÂL</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>QUILT</td>
<td>NC</td>
<td>1</td>
<td>GLONŢ</td>
<td>1</td>
<td>NC</td>
<td>NOTA</td>
<td>1</td>
<td>NC</td>
</tr>
<tr>
<td>V</td>
<td>SHUDDER</td>
<td>NC</td>
<td>2</td>
<td>ABTINE</td>
<td>3</td>
<td>NC</td>
<td>ESKÜSZIK</td>
<td>3</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>AZALEA</td>
<td>RO - HU</td>
<td>3</td>
<td>PĂGĂN</td>
<td>2</td>
<td>EN - HU</td>
<td>PORTRÉ</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>STEALTH</td>
<td>NC</td>
<td>1</td>
<td>BRÂU</td>
<td>1</td>
<td>NC</td>
<td>RONGY</td>
<td>1</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>DEMOGRAPHY</td>
<td>RO - HU</td>
<td>4</td>
<td>PARADÂ</td>
<td>3</td>
<td>EN - HU</td>
<td>NACIONALIZMUS</td>
<td>6</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>ERRATIC</td>
<td>RO</td>
<td>3</td>
<td>AUJSTAT</td>
<td>3</td>
<td>EN</td>
<td>VITÉZ</td>
<td>2</td>
<td>RO</td>
</tr>
<tr>
<td>N</td>
<td>MARROW</td>
<td>NC</td>
<td>2</td>
<td>MOFT</td>
<td>1</td>
<td>NC</td>
<td>OSTROM</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>MUMBLE</td>
<td>NC</td>
<td>2</td>
<td>DEGAJA</td>
<td>3</td>
<td>NC</td>
<td>TÖPRENG</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>PALETTE</td>
<td>RO - HU</td>
<td>2</td>
<td>MONOPOL</td>
<td>3</td>
<td>EN - HU</td>
<td>SZPONZOR</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>LOCUST</td>
<td>RO</td>
<td>2</td>
<td>TRAJECTORIE</td>
<td>5</td>
<td>EN</td>
<td>KASZA</td>
<td>2</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>N</td>
<td>HALLMARK</td>
<td>NC</td>
<td>2</td>
<td>BĂLĂRIE</td>
<td>4</td>
<td>NC</td>
<td>ÖNBIZALOM</td>
<td>4</td>
<td>NC</td>
</tr>
<tr>
<td>N</td>
<td>WHIM</td>
<td>NC</td>
<td>1</td>
<td>PUDOARE</td>
<td>4</td>
<td>NC</td>
<td>KOZMA</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>OCTOPUS</td>
<td>RO - HU</td>
<td>3</td>
<td>RIVAL</td>
<td>2</td>
<td>EN - HU</td>
<td>MONITOR</td>
<td>3</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>WEIR</td>
<td>NC</td>
<td>1</td>
<td>DOVLEAC</td>
<td>2</td>
<td>NC</td>
<td>KÖZMÛ</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>REGENT</td>
<td>RO - HU</td>
<td>2</td>
<td>NATURALIST</td>
<td>4</td>
<td>EN - HU</td>
<td>PERSPEKTÍVA</td>
<td>4</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>EGALITARIAN</td>
<td>RO</td>
<td>6</td>
<td>CONCIS</td>
<td>2</td>
<td>EN</td>
<td>PUSKÁS</td>
<td>2</td>
<td>RO</td>
</tr>
<tr>
<td>N</td>
<td>CRANNY</td>
<td>NC</td>
<td>1</td>
<td>TUNET</td>
<td>2</td>
<td>NC</td>
<td>UGAR</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>LECTERN</td>
<td>NC</td>
<td>2</td>
<td>GUDRON</td>
<td>2</td>
<td>NC</td>
<td>FICKÓ</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>PEASANTRY</td>
<td>NC</td>
<td>3</td>
<td>CRĂPĂTURĂ</td>
<td>4</td>
<td>NC</td>
<td>RANGADÓ</td>
<td>3</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>PIGTAIL</td>
<td>NC</td>
<td>2</td>
<td>AMANT</td>
<td>2</td>
<td>NC</td>
<td>FÁTYOL</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>EXCRETE</td>
<td>RO</td>
<td>2</td>
<td>REITERA</td>
<td>4</td>
<td>EN</td>
<td>TOLMÁCSOL</td>
<td>3</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>PUMA</td>
<td>RO – HU</td>
<td>2</td>
<td>STRATEG</td>
<td>2</td>
<td>EN - HU</td>
<td>PIRAMIS</td>
<td>3</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>HESSIAN</td>
<td>RO</td>
<td>3</td>
<td>INSURECȚIE</td>
<td>5</td>
<td>EN</td>
<td>JOBBÁGY</td>
<td>2</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>COUNTERCLAIM</td>
<td>NC</td>
<td>3</td>
<td>COCHILIE</td>
<td>4</td>
<td>NC</td>
<td>BONYODALOM</td>
<td>4</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>EMIR</td>
<td>RO - HU</td>
<td>2</td>
<td>STEREOTIP</td>
<td>4</td>
<td>EN - HU</td>
<td>INTENZITÁS</td>
<td>4</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>SPLEEN</td>
<td>RO</td>
<td>1</td>
<td>SANCTUAR</td>
<td>3</td>
<td>EN</td>
<td>ÁBRÁZAT</td>
<td>3</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>ALUM</td>
<td>RO</td>
<td>2</td>
<td>PERCHEZITIE</td>
<td>5</td>
<td>EN</td>
<td>SAKK</td>
<td>1</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>COVEN</td>
<td>NC</td>
<td>2</td>
<td>CLEPSIDRÁ</td>
<td>3</td>
<td>NC</td>
<td>MAGÁNÚGY</td>
<td>3</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>SOLLIOQUY</td>
<td>RO – HU</td>
<td>4</td>
<td>VARIAȚIE</td>
<td>5</td>
<td>EN - HU</td>
<td>TERMINUS</td>
<td>3</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>REFECTORY</td>
<td>RO – HU</td>
<td>4</td>
<td>COLAJ</td>
<td>2</td>
<td>EN - HU</td>
<td>TÉZIS</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>UBIQUITOUS</td>
<td>RO</td>
<td>4</td>
<td>CAUSTIC</td>
<td>3</td>
<td>EN</td>
<td>PIARISTA</td>
<td>4</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>COMMUNIQUE</td>
<td>RO – HU</td>
<td>4</td>
<td>DISERTAȚIE</td>
<td>5</td>
<td>EN - HU</td>
<td>DEMAGÓGIA</td>
<td>5</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>ATOLL</td>
<td>RO - HU</td>
<td>2</td>
<td>NIHILISM</td>
<td>3</td>
<td>EN - HU</td>
<td>KONZUL</td>
<td>2</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>JOVIAL</td>
<td>RO</td>
<td>3</td>
<td>LAMENTABIL</td>
<td>4</td>
<td>EN</td>
<td>KARITATÍV</td>
<td>4</td>
<td>EN - RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>SKYLARK</td>
<td>NC</td>
<td>2</td>
<td>PURICE</td>
<td>3</td>
<td>NC</td>
<td>ZSOLTÁR</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>GAUCHE</td>
<td>NC</td>
<td>1</td>
<td>DESTOINIC</td>
<td>3</td>
<td>NC</td>
<td>HITVÁNY</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>AUGUR</td>
<td>RO</td>
<td>2</td>
<td>DILATA</td>
<td>3</td>
<td>EN</td>
<td>REKLÁMOZ</td>
<td>3</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>ERYTHROCYTE</td>
<td>RO</td>
<td>4</td>
<td>LAMELÁ</td>
<td>3</td>
<td>EN</td>
<td>CSIZMADIA</td>
<td>4</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>MARSUPIAL</td>
<td>RO</td>
<td>4</td>
<td>FILIGRAN</td>
<td>3</td>
<td>EN</td>
<td>BICSKA</td>
<td>2</td>
<td>RO</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>BAWDY</td>
<td>NC</td>
<td>2</td>
<td>BUIMAC</td>
<td>2</td>
<td>NC</td>
<td>KOPÁR</td>
<td>2</td>
<td>NC</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 7 The Questionnaire

The following questions will help me understand more of your language background. Please read the questions carefully and answer them honestly. There are no right or wrong answers.

1. At what age did you start learning each of these languages?
   a. Hungarian ______
   b. Romanian ______
   c. English ______
   d. Other(s) (please name) _______________ ______

2. On a scale of 1 to 5 please rate how proficient you feel using each language:
   a. Hungarian
      1 2 3 4 5
      Not proficient Fairly proficient Very proficient
   b. Romanian
      1 2 3 4 5
      Not proficient Fairly proficient Very proficient
   c. English
      1 2 3 4 5
      Not proficient Fairly proficient Very proficient

3. On a scale of 1 to 5, please rate (by circling the number) how confident you feel using each language in everyday situations (e.g. in a restaurant, shopping, chatting to friends).
   a. Hungarian
      1 2 3 4 5
      Not confident Fairly confident Very confident
   b. Romanian
      1 2 3 4 5
      Not confident Fairly confident Very confident
c. English

Not confident | Fairly confident | Very confident

4. Please rate how confident you feel using each language in a *formal setting* (e.g. job interview, presentation, business meeting)

a. Hungarian

Not confident | Fairly confident | Very confident

b. Romanian

Not confident | Fairly confident | Very confident

c. English

Not confident | Fairly confident | Very confident

5. Imagine you work at an international company and are responsible for *emails, phone calls, meetings* etc. Please rate each option. Think of professionalism or efficiency, not preference.

a. Hungarian

Not proficient | Fairly proficient | Very proficient

b. Romanian

Not proficient | Fairly proficient | Very proficient

c. English

Not proficient | Fairly proficient | Very proficient
6. Imagine you moved to an **English-speaking country**. Please rate how confident would you feel in using **English** in everyday situations (e.g. calling locals, ordering food, going on a date with a local)

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Not confident  Fairly confident  Very confident

7. You are writing an essay in English. The right word/expression is just not coming to your mind. Which of the following describes you best? In most cases:
   a. I think of suitable words in **Hungarian** and translate them
   b. I think of suitable words in **Romanian** and see if they work in English or translate them
   c. Other strategy (describe):

8. You are reading in English. You find an unknown word/expression. Which of the following describes you best? In most cases:
   a. I try to find a **Hungarian** translation of the word
   b. I try to think if I’ve seen the word in **Romanian**
   c. I try to find a **Romanian** translation of the word
   d. I try to guess it from the context
   e. Other strategy (describe)

9. Thinking of your **Romanian** vocabulary, how much would you say it helps you in learning or functioning (speaking/writing) in **English**?
   a. It definitely helps
   b. In many cases it can be useful
   c. Sometimes it helps, sometimes it doesn’t
   d. It doesn’t make a difference at all
   e. It can be confusing and misleading

10. If you received the handout on cognates before the study, please answer the question below. If not, please turn over 😊

How much did the task/handout on cognates help in making a connection between English and Romanian? How much do you think it would help / would have helped if you received similar instructions more often when learning English? How important/useful you think learning English through Romanian instead of Hungarian would be? Please comment in a couple of sentences (approximately 100 words).
**Appendix 8**

**Cognate Task**

We often think that learning a foreign language is a tremendous task. Truth to be told, a good number of languages, especially Romance languages, such Spanish and Italian, have a lot in common. More than we think. It is estimated that approximately 25% of an English text will be familiar to Romance language speakers, and this includes Romanian.

To test this, please underline all the common words in the three texts.

Un grup de animale - un leu, o zebră, o girafă, și un hipopotam - evadează de la o grădină zoologică din New York și sosesc pe o insulă unde vor avea multe aventuri cu un trib de lemuri și cu amicii ei pinguini foarte inteligenti.

Un grupo de animales - un León, una Cebra, una Jirafa, y una Hipopótamo- escapan de un zoológico en New York y arriban a una isla donde tendrán muchas aventuras con una tribu de lémures y con sus amigos los inteligentes pingüinos.

A group of animals - a lion, a zebra, a giraffe and a hippopotamus - escape from a zoo in New York and arrive to an island where they will have many adventures with a tribe of lemurs and their very intelligent penguin friends.
Appendix 9 - Print-screens from Qualtrics.com (Study Two)

On a scale of 0 to 5 please rate how proficient you feel using each language:

NOT proficient  1  2  3  4  VERY proficient

HUNGARIAN

ROMANIAN

ENGLISH

How confident do you feel using each language in EVERYDAY SITUATIONS (e.g. in a restaurant, shopping, chatting to friends)?

These page timer metrics will not be displayed to the recipient:

First Click  0 seconds
Last Click  0 seconds
Page Submit  0 seconds
Click Count  0 clicks

19

POOR: We are poor

a. have no money
b. feel happy

c. are very interested
d. do not like to work hard

*The full test battery used in Study Two is available upon request