



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

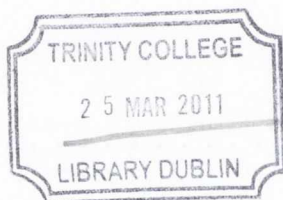
**The relationship between
phonological short-term memory
and the development of L2 vocabulary
and collocational knowledge:**

A longitudinal study of adult Polish learners of English
at A2 and B1 levels of English proficiency

Agnieszka Skrzypek
Ph.D. in Applied Linguistics

Thesis submitted for the Degree of Doctor of Philosophy
Centre for Language and Communication Studies
University of Dublin
Trinity College

2010



THOS
9012



© Agnieszka Skrzypek

© Agnieszka Skrzypek

“Memory is the primary and fundamental power, without which
there could be no other intellectual operation.”

Samuel Johnson, Idler No. 44
February 17, 1759

DECLARATION

I hereby declare that this thesis submitted in candidature for the degree of Doctor of Philosophy at Trinity College Dublin is entirely my own work and has not been previously submitted for a degree at this or any other university.

I agree that the Library of Trinity College Dublin may lend or copy the thesis upon request.

Abstract

The present study examines the relationship between individual differences in Phonological Short-term Memory (PSTM) and L2 vocabulary/collocation learning in adult L2 learners in a longitudinal classroom-based study at the A2 and B1 level of L2 proficiency. PSTM is one of the four components of the Working Memory model (Baddeley, 1986; Baddeley, 2000; Baddeley, 2003; Baddeley, 2007; Baddeley & Hitch, 1974). It is theorised to comprise two sub-components: the nonarticulatory phonological store which is responsible for holding phonological information for up to a few seconds, and the articulatory rehearsal mechanism which refreshes memory traces held in the phonological store and prevents them from decaying. The Working Memory model has received a considerable amount of attention from academics working the area of cognition and language. Even though PSTM is seen as limited in capacity to a few memory items, it has generated an extensive amount of research over the last 30 years, particularly in developmental studies on children and laboratory-based experiments involving adults. PSTM has been shown to be linked significantly with L1 vocabulary knowledge/learning in children, but studies involving adults carried out to date present a less conclusive picture.

The organization of the thesis is as follows. Chapter 1 presents a number of crucial features which have not been assembled in a single study in PSTM research with respect to SLA in adult learners and that are addressed in the present work. It outlines our six research questions, along with an account of the relevant research context.

Chapter 2 discusses developments in the evolving concept of working memory that contributed to the formulation of the Baddeley and Hitch Working Memory model. It provides an overview of the current evidence for the existence of PSTM, presenting the PSTM account (Baddeley, Gathercole & Papagno, 1998) and the linguistic account (Snowling, Chiat & Hulme, 1991), which propose opposite directions of causality in the relationship between PSTM and vocabulary knowledge/learning. Since the Baddeley and Hitch Working Memory model does not directly address the mechanisms responsible for the interaction between PSTM and LTM, Chapter 2 presents a number of models that simulate PSTM computationally. This chapter provides a brief account of functions of PSTM documented by research to date and gives an account of two PSTM tasks that have been used in the current study, i.e. Serial Nonword Recall and Serial Nonword Recognition.

Chapter 3 is divided into two main thematic sections. The first thematic section introduces different approaches to defining a word and to the organization of the bilingual mental lexicon. Since one of our L2 vocabulary measures was designed with reference to frequency lists, the frequency factor in vocabulary learning is briefly discussed. The second thematic section takes a closer look at the relationship between PSTM and L2 vocabulary knowledge/learning. It focuses on the issue of language proficiency and strength of relationship between PSTM and vocabulary learning, and the causal direction of the relationship. It further explores the premises upon which the PSTM account and the linguistic account are based.

Chapter 4 is also divided into two main thematic sections. The first thematic section contains an overview of two major approaches to collocations, i.e. the frequency-based approach and the phraseological approach. It discusses the theoretical relevance of collocations in language learning, and reviews models of L1 and L2 collocational learning in the context of formula-based approaches to language learning. It also presents test formats used to tap L2 collocation knowledge and statistical procedures used to calculate the frequency and strength of collocations. The second thematic section explores the link between PSTM and L2 collocation learning. It presents one laboratory-based study (Ellis & Sinclair, 1996) indicating that PSTM may be linked not only to L2 vocabulary but also L2 collocation learning in adult learners at the onset of L2 learning. In addition, some indirect classroom-based evidence for patterns expected at a post-beginning stage of L2 learning is reviewed.

Chapters 5 and 6 contain a detailed description of our research instruments and our pilot study, while Chapters 7 and 8 present our data analysis and conclusions. The findings from our study indicate that PSTM is significantly linked with L2 vocabulary and L2 collocation knowledge at the A2 level of L2 proficiency. At the B1 level the link between the PSTM and L2 vocabulary becomes insignificant when Exposure is partialled out, suggesting a diminishing influence of PSTM on L2 vocabulary knowledge/learning. The relationship between PSTM and L2 collocations at the B1 level is significant, when PSTM is tapped by Serial Nonword Recall. However, when PSTM is indexed by Serial Nonword Recognition, the link between PSTM and L2 collocations is nonsignificant. Possible causes for that conflicting evidence at the B1 level are discussed. Additionally, a cross-lagged correlational analysis is presented that indicates that PSTM (tapped by Serial Nonword Recall) is a causal determinant of L2 vocabulary/collocation knowledge at the A2 level of proficiency.

Acknowledgements

I would like to thank my supervisor Dr John Harris for all his encouragement, support and guidance. I am also indebted to him for his tireless attention to the detail of my project and of my writing.

I would also like to thank Prof David Singleton for his confidence in my ability to complete this research.

I would also like to express gratitude to Dr Elisabet Service for her assistance with the devising of the PSTM instruments I have deployed in this study.

Finally, I wish to acknowledge the support of the IRCHSS for the broader project of which this Ph.D. project forms a part.

Table of Contents

List of Tables	vii
List of Figures	x
List of Equations	xii
CHAPTER 1: INTRODUCTION	1
1.1. Research questions	3
1.2. Operational definitions	10
1.2.1. Phonological Short-term Memory (PSTM)	10
1.2.2. L2 vocabulary knowledge and learning	11
1.2.3. L2 collocation knowledge and learning	12
1.2.4. L1 vocabulary knowledge	14
1.2.5. Exposure to L2 outside the classroom	14
CHAPTER 2: PHONOLOGICAL SHORT-TERM MEMORY (PSTM) AND ADULT SECOND LANGUAGE LEARNING	16
2.1. Introduction	16
2.2. The evolution of the concept of Working Memory and Short-term Memory	17
2.3. The Working Memory model of Baddeley and Hitch	20
2.3.1. The Architecture of the Working Memory model	22
2.3.2. Current evidence for PSTM	24
Phonological similarity effect	25
Word length effect and articulation rate	26
Articulatory suppression	26
The influence of long-term knowledge	27
2.3.3. Limitations of the Baddeley and Hitch Working Memory model – Nairne's critique	29
2.4. Mechanisms for short-term and long-term learning in computational STM models	30
2.4.1. Approaches to the storage and retrieval of serial order	31
2.4.2. Word learning mechanisms in selected computational STM models	33
The Burgess and Hitch (1999) model	33

The Hartley and Houghton (1996) model	36
The Gupta and MacWhinney (1997) model	38
2.5. Brief overview of the functions of PSTM	43
2.5.1. PSTM, L2 vocabulary and collocations, and the level of L2 proficiency	43
2.5.2. The phonological store and the articulatory rehearsal component in L2 learning	48
2.5.3. The direction of causation in the link between PSTM and knowledge of L2 vocabulary and collocations	49
2.6. Selected PSTM measures: serial recall and recognition of nonwords	51
2.7. PSTM scores and the language of testing	53
 CHAPTER 3: VOCABULARY BREADTH AND MEMORY IN ADULT SECOND LANGUAGE LEARNERS	 55
3.1. Introduction	55
3.2. Defining vocabulary and word knowledge	56
3.2.1. Approaches to defining the word	56
3.2.2. Breadth, depth and other aspects of word knowledge	58
3.2.3. The bilingual mental lexicon	61
3.3. Frequency and vocabulary learning	63
3.3.1. Differences in word frequency	64
3.3.2. English language corpora and frequency lists	66
English language corpora	66
Lemmas and word families as counting units	69
English language frequency lists	71
3.3.3. The frequency model and vocabulary learning	73
3.4. Measuring vocabulary breadth	76
3.4.1. The receptive versus productive distinction in vocabulary breadth	77
3.4.2. Receptive measures of vocabulary breadth	78
3.4.3. Productive measures of vocabulary breadth	82
3.4.4. Advantages and disadvantages of selected formats of vocabulary breadth tests	84
3.5. Vocabulary breadth and the Common European Framework	88
3.6. Phonological Short-term Memory (PSTM) and L2 vocabulary knowledge/learning	91
3.6.1. Early studies on PSTM capacity and vocabulary learning	91

3.6.2. Language proficiency and strength of relationship between PSTM and vocabulary knowledge/learning _____	92
3.6.3. The causal direction of relationship between PSTM and vocabulary knowledge/learning _____	95

CHAPTER 4: COLLOCATIONS AND MEMORY IN ADULT SECOND LANGUAGE

LEARNERS _____ 101

4.1. Introduction _____ 101

4.2. The notion of collocation _____ 102

4.2.1. Defining and identifying collocations _____	102
4.2.2. Semantic prosody and near-synonymy _____	105
4.2.3. Principles of interpretation and economy of effort _____	106
The idiom principle and the open choice principle _____	106
The economy of effort principle _____	108

4.3. Collocations and formulaic language: terminology, approaches and functions _____ 109

4.3.1. Collocations and formulaic language _____	110
4.3.2. Approaches to studying collocations and formulaic language _____	111
The frequency-based approach _____	112
The phraseological approach _____	114
4.3.3. Selected functions of collocations and formulaic language _____	116

4.4. Collocations and formulas in the theory of language learning _____ 118

4.4.1. L1 collocation learning _____	119
The Ellis (1996) model of L1 collocation learning _____	119
Formulas in child L1 learning _____	120
4.4.2. L2 collocation learning _____	122
The Wray (2002) model of L2 collocation learning _____	123
Formulas in adult L2 learning _____	124

4.5. Measurement of collocation knowledge _____ 126

4.5.1. Receptive and productive tests of collocation knowledge _____	126
4.5.2. Statistical measurements of collocational links between words _____	131
Raw frequencies of collocations _____	131
Measures of collocational strength and certainty _____	132

4.6. Phonological Short-term Memory and adult L2 collocation learning _____ 134

4.6.1. PSTM and collocation in adult L2 learners: experimental evidence _____	135
4.6.2. PSTM and collocation in adult classroom L2 learners _____	137

CHAPTER 5: METHODOLOGY AND PILOTING	139
5.1. Introduction	139
5.2. Research design	139
5.2.1. Design of the main study	139
Dependent and independent variables	139
The administration of the test battery	141
Research questions	142
5.2.2. Objectives of the pilot study	143
5.3. Piloting of the instruments	144
5.3.1. Serial nonword tasks	144
The process of constructing experimental nonwords	144
Normality and distribution of scores	146
5.3.2. Gitsaki Collocation Tests	150
Minor alterations to Gitsaki Collocation Tests	151
Normality and distribution of scores	152
Item difficulty index	153
5.3.3. Vocabulary Levels Test	155
Parallel versions of the Vocabulary Levels Test	156
The choice of frequency levels	158
Minor alterations to the Vocabulary Levels Test	159
Creation of the productive 1,000 Vocabulary Levels Test	163
Comparing the amended versions of the Vocabulary Levels Test	163
5.3.4. L1 Vocabulary Test	169
Normality and distribution of scores	170
Item difficulty index	172
CHAPTER 6: SAMPLE AND RESEARCH INSTRUMENTS	173
6.1. Introduction	173
6.2. Sample and language course description	173
6.2.1. Polish migration to Ireland prior to 2008	173
6.2.2. Sample description	174
6.2.3. English language course description	176
6.3. Instruments	176
6.3.1. Background questionnaire	176

6.3.2. Serial Nonword Recall	177
6.3.3. Serial Nonword Recognition	180
6.3.4. Gitsaki Collocation Tests	182
6.3.5. Vocabulary Levels Test	185
General information	185
1,000-word frequency level tests	186
2,000 and 3,000-word frequency levels tests	187
6.3.6. Progress Vocabulary and Collocation Tests	189
6.3.7. L1 Vocabulary Test	191
6.3.8. OUP Quick Placement Test	195
CHAPTER 7: RESULTS AND DISCUSSION	196
7.1. Introduction	196
7.2. Brief overview of measures and research questions	197
7.3. Distribution of scores and assumptions of data	199
Serial Nonword Recall and Recognition scores (Time 1, Time 2)	200
L1 Vocabulary Test scores (Time 1)	203
Exposure to L2 outside the classroom (between Time 1 and Time 2)	204
A2 and B1 Vocabulary Levels Test scores (Time 1, Time 2)	205
A2 and B1 Gitsaki Collocation Test scores (Time 1, Time 2)	208
A2 and B1 Progress Vocabulary Test scores (Time 2)	210
A2 and B1 Progress Collocation Test scores (Time 2)	211
7.4. Reliability and validity of the research instruments	212
7.5. Correlations among main variables	214
Correlations among independent variables and L2 vocabulary scores in the A2 group	215
Correlations among independent variables and L2 vocabulary scores in the B1 group	221
Correlations among independent variables and L2 collocation scores in the A2 group	226
Correlations among independent variables and L2 collocation scores in the B1 group	229
7.6. Regression models of the link between PSTM and L2 vocabulary	232
PSTM and L2 vocabulary scores in the A2 group	232
PSTM and L2 vocabulary scores in the B1 group	238
7.7. Regression models of the link between PSTM and L2 collocations	241
PSTM scores and L2 collocation scores in the A2 group	241

PSTM scores and L2 collocation scores in the B1 group	244
7.8. Crossed-lagged correlational analyses of the link between PSTM and L2 vocabulary/collocations	246
CHAPTER 8: CONCLUSIONS	254
8.1. A summary of the current research findings	254
8.1.1. PSTM capacity and L2 vocabulary/collocation knowledge and learning in A2 and B1 adult L2 learners	254
8.1.2. Sub-components of PSTM, L2 vocabulary/collocations and adult L2 proficiency	263
8.1.3. PSTM and adult L2 vocabulary/collocation: causal direction and proficiency	267
8.2. Limitations of the present study	272
8.2.1. Limitations of a classroom-based longitudinal design	272
8.2.2. Limitations of the Serial Nonword Recall and the Serial Nonword Recognition tasks	273
8.3. Suggestions for future research	275
8.3.1. The impact of phonological proximity between L1 and L2 on the relationship between PSTM and L2 learning	275
8.3.2. Types of nonwords used in PSTM tasks	276
8.3.3. Serial Nonword Recognition and the construct it measures in adults	277
8.3.4. PSTM and the second language classroom	278
8.3.5. PSTM and free productive use of L2 vocabulary/collocations in adults	278
8.3.6. Construction of new L2 collocation tests	279
8.4. Conclusion	279
APPENDIX A: Consent form	281
APPENDIX B: Background questionnaire	283
APPENDIX C: Serial Nonword Recall task	288
APPENDIX D: Serial Nonword Recognition task	290
APPENDIX E: A2 and B1 Gitsaki Collocation Tests	292
APPENDIX F: Vocabulary Levels Tests	306
APPENDIX G: Progress Vocabulary and Collocation Test	323
APPENDIX H: Polish Vocabulary Test	343
APPENDIX I: OUP Quick Placement Test	359
APPENDIX J: Results of the pilot study	362
APPENDIX K: Results of the main study	367
REFERENCES	374

List of Tables

TABLE 3.1. TYPES OF WORD KNOWLEDGE ACCORDING TO NATION (2001b: 27)	60
TABLE 3.2. EXAMPLES OF WORDS FROM THE FIRST, SECOND AND FIFTH 1,000 FREQUENCY BANDS AND THEIR OCCURRENCES IN THE BNC (FROM KILGARRIFF, 2006)	65
TABLE 3.3. THE CEFR OVERALL PROFICIENCY DESCRIPTORS FOR THE A2 AND B1 LEVELS	89
TABLE 3.4. THE CEFR LEVELS AND AVERAGE EFL VOCABULARY SCORES	90
TABLE 4.1. SELECTED COLLOCATION TYPES IN THE BBI TYPOLOGY (BENSON ET AL., 1986)	116
TABLE 5.1. OUTLINE OF DEPENDENT AND INDEPENDENT VARIABLES	140
TABLE 5.2. TESTING BATTERY AND TIMES OF TEST ADMINISTRATION	141
TABLE 5.3. PILOTING – EXAMPLES OF HIGH-WORDLIKE NONWORDS AND SIMILAR-SOUNDING POLISH WORDS	145
TABLE 5.4. PILOTING – DISTRIBUTION OF SCORES IN SERIAL NONWORD TASKS (N=20)	148
TABLE 5.5. PILOTING – DISTRIBUTION OF SCORES IN GITSAKI COLLOCATION TESTS (N=15)	153
TABLE 5.6. PILOTING – COLLOCATIONS WITH THE HIGHEST AND LOWEST ITEM DIFFICULTY COEFFICIENTS	155
TABLE 5.7. PILOTING – ITEM DIFFICULTY COEFFICIENTS OF COGNATES FROM THE 2,000 AND 3,000 VOCABULARY LEVELS TEST (N=15, LEVEL=A2)	162
TABLE 5.8. EXAMPLES OF WORDS INCLUDED IN THE 2,000 AND 3,000 VOCABULARY LEVELS TEST AND THEIR MODE OF TESTING	164
TABLE 5.9. EXAMPLES OF IDENTICAL SEQUENCING OF SOME ITEMS IN THE RECEPTIVE AND PRODUCTIVE 2,000 AND 3,000 VOCABULARY LEVELS TEST	165
TABLE 5.10. PILOTING - DESCRIPTIVE STATISTICS FOR PARALLEL VERSIONS OF THE VOCABULARY LEVELS TEST (N=15, LEVEL=A2)	166
TABLE 5.11. PILOTING - DESCRIPTIVE STATISTICS FOR PARALLEL VERSIONS OF THE VOCABULARY LEVELS TEST (N=15, LEVEL=B1)	167
TABLE 5.12. PILOTING - ITEM DIFFICULTY COEFFICIENTS OF SELECTED ITEMS FROM THE RECEPTIVE 1,000 VOCABULARY LEVELS TEST	169
TABLE 5.13. PILOTING – DESCRIPTIVE STATISTICS FOR L1 VOCABULARY TEST SCORES (N=13)	171
TABLE 5.14. PILOTING – L1 VOCABULARY TEST ITEMS WITH AN ITEM DIFFICULTY COEFFICIENT BELOW 0.7 (N=13)	172
TABLE 6.1. BACKGROUND INFORMATION ABOUT PARTICIPANTS	175
TABLE 6.2. WORDLIKENESS RATINGS FOR NONWORDS IN THE SERIAL NONWORD RECALL TASK	178
TABLE 6.3. FREQUENCY OF NONWORD OCCURRENCE IN 2-, 3- AND 4-ITEM LISTS IN THE SERIAL RECALL TASK	179
TABLE 6.4. WORDLIKENESS RATINGS FOR NONWORDS IN THE SERIAL NONWORD RECOGNITION TASK	181
TABLE 6.5. FREQUENCY OF NONWORD USE IN 4-, 5- AND 6-ITEM LISTS IN SERIAL NONWORD RECOGNITION	182
TABLE 6.6. THE BBI TYPES AND BNC FREQUENCIES OF OCCURRENCE OF ENGLISH COLLOCATIONS IN GITSAKI TESTS	184
TABLE 6.7. EXAMPLES OF WORDS TESTED IN THE A2/B1 PROGRESS VOCABULARY TESTS AND THEIR MODE OF TESTING ...	190
TABLE 6.8. THE BBI TYPES OF COLLOCATIONS TESTED IN THE A2/B1 PROGRESS COLLOCATION TESTS	191
TABLE 6.9. FREQUENCIES OF OCCURRENCE OF INFLECTED FORMS OF A SELECTED POLISH VERB IN THE IPI PAN CORPUS ...	193

TABLE 6.10. FREQUENCIES OF OCCURRENCE OF INFLECTED FORMS OF A SELECTED NOUN AND ADJECTIVE IN THE IPI PAN CORPUS.....	194
TABLE 7.1. OVERVIEW OF VARIABLES, MEASURES AND SCORING METHODS	197
TABLE 7.2. OVERVIEW OF RESEARCH QUESTIONS AND SECTIONS WHERE THE QUESTIONS ARE ADDRESSED	198
TABLE 7.3. RELIABILITY COEFFICIENTS OF THE RESEARCH INSTRUMENTS.....	213
TABLE 7.4. SIMPLE INTERCORRELATIONS AMONG INDEPENDENT VARIABLES AND A2 VOCABULARY MEASURES (N=30)	216
TABLE 7.5. SIMPLE INTERCORRELATIONS AMONG INDEPENDENT VARIABLES AND B1 VOCABULARY MEASURES (N=30)	222
TABLE 7.6. SIMPLE INTERCORRELATIONS AMONG INDEPENDENT VARIABLES AND A2 COLLOCATION MEASURES (N=30)	227
TABLE 7.7. SIMPLE INTERCORRELATIONS AMONG INDEPENDENT VARIABLES AND B1 COLLOCATION MEASURES (N=30)	230
TABLE 7.8. SUMMARY OF HIERARCHICAL MULTIPLE REGRESSIONS FOR VARIABLES PREDICTING GAINS IN A2 VOCABULARY LEVELS TEST GAIN SCORES (N=30)	234
TABLE 7.9. SUMMARY OF HIERARCHICAL MULTIPLE REGRESSIONS FOR VARIABLES PREDICTING THE A2 VOCABULARY LEVELS TEST SCORES AT TIME 2 (N=30)	237
TABLE 7.10. SUMMARY OF HIERARCHICAL MULTIPLE REGRESSIONS FOR VARIABLES PREDICTING GAINS IN B1 VOCABULARY LEVELS TEST SCORES (N=30)	239
TABLE 7.11. SUMMARY OF HIERARCHICAL MULTIPLE REGRESSIONS FOR VARIABLES PREDICTING THE B1 VOCABULARY LEVELS TEST SCORES AT TIME 2 (N=30)	240
TABLE 7.12. SUMMARY OF HIERARCHICAL MULTIPLE REGRESSIONS FOR VARIABLES PREDICTING GAINS IN A2 GITSAKI COLLOCATION TEST GAIN SCORES (N=30)	242
TABLE 7.13. SUMMARY OF HIERARCHICAL MULTIPLE REGRESSIONS FOR VARIABLES PREDICTING A2 GITSAKI COLLOCATION TEST SCORES AT TIME 2 (N=30)	244
TABLE 7.14. SUMMARY OF HIERARCHICAL MULTIPLE REGRESSIONS FOR VARIABLES PREDICTING B1 GITSAKI COLLOCATION TEST SCORES AT TIME 2 AND GAIN SCORES (N=30)	245
TABLE A.1. THE BBI TYPES AND BNC FREQUENCIES OF OCCURRENCE OF ENGLISH COLLOCATIONS TESTED IN THE A2 GITSAKI COLLOCATION TEST	297
TABLE A.2. THE BBI TYPES AND BNC FREQUENCIES OF OCCURRENCE OF ENGLISH COLLOCATIONS TESTED IN THE B1 GITSAKI COLLOCATION TEST	300
TABLE A.3. THE BBI TYPES AND BNC FREQUENCIES OF OCCURRENCE OF ENGLISH COLLOCATIONS TESTED IN BOTH A2 AND B1 GITSAKI COLLOCATION TESTS	303
TABLE A.4. WORDS TESTED RECEPTIVELY AND PRODUCTIVELY IN VERSIONS 1 & 2 OF 2,000 AND 3,000 VOCABULARY LEVELS TESTS	320
TABLE A.5. WORDS TESTED ONLY RECEPTIVELY IN VERSION 1 & 2 OF 2,000 AND 3,000 VOCABULARY LEVELS TESTS	321
TABLE A.6. WORDS TESTED ONLY PRODUCTIVELY IN VERSION 1 & 2 OF 2,000 AND 3,000 VOCABULARY LEVELS TESTS ..	322
TABLE A.7. WORDS TESTED IN THE A2 AND B1 PROGRESS VOCABULARY TESTS AND THEIR MODE OF TESTING.....	338
TABLE A.8. THE BBI TYPES OF COLLOCATIONS TESTED IN THE A2 PROGRESS COLLOCATION TEST	339
TABLE A.9. THE BBI TYPES OF COLLOCATIONS TESTED IN THE B1 PROGRESS COLLOCATION TEST	340
TABLE A.10. ENGLISH TRANSLATIONS OF POLISH WORDS FROM THE POLISH VOCABULARY TEST	346
TABLE A.11. FREQUENCIES OF OCCURRENCE OF TARGET POLISH WORDS IN THE IPI PAN CORPUS	348

TABLE A.12. FREQUENCIES OF OCCURRENCE OF TARGET POLISH VERBS AND THEIR INFLECTED FORMS IN THE IPI PAN CORPUS	
— ALPHABETICAL ORDER	349
TABLE A.13. FREQUENCIES OF OCCURRENCE OF TARGET POLISH NOUNS AND THEIR INFLECTED FORMS IN THE IPI PAN	
CORPUS — ALPHABETICAL ORDER	353
TABLE A.14. FREQUENCIES OF OCCURRENCE OF TARGET POLISH ADJECTIVES AND THEIR INFLECTED FORMS IN THE IPI PAN	
CORPUS — ALPHABETICAL ORDER	357
TABLE A.15. PILOTING — WORDLIKENESS RATINGS AND PRONUNCIATION DIFFICULTY OF 40 LOW-WORDLIKE NONWORDS	362
TABLE A.16. SERIAL NONWORD RECALL AND RECOGNITION SCORES IN A2 AND B1 GROUPS (TIME 1)	367
TABLE A.17. SERIAL NONWORD RECALL SCORES IN HIGH AND LOW SERIAL NONWORD RECALL GROUPS (TIME 1).....	367
TABLE A.18. SERIAL NONWORD RECOGNITION SCORES IN HIGH AND LOW SERIAL NONWORD RECOGNITION GROUPS (TIME 1).....	367
TABLE A.19. SERIAL NONWORD RECALL AND RECOGNITION SCORES IN A2 AND B1 GROUPS (TIME 2)	368
TABLE A.20. POLISH VOCABULARY TEST SCORES IN A2 AND B1 GROUPS (TIME 1)	368
TABLE A.21. EXPOSURE SCORES IN A2 AND B1 GROUPS	368
TABLE A.22. A2 VOCABULARY LEVELS TEST SCORES (TIME 1 AND TIME 2)	368
TABLE A.23. B1 VOCABULARY LEVELS TEST SCORES (TIME 1 AND TIME 2).....	369
TABLE A.24. A2 AND B1 VOCABULARY LEVELS TEST GAIN SCORES	369
TABLE A.25. A2 GITSAKI COLLOCATION TEST SCORES (TIME 1 AND TIME 2)	369
TABLE A.26. B1 GITSAKI COLLOCATION TEST SCORES (TIME 1 AND TIME 2)	369
TABLE A.27. A2 AND B1 GITSAKI COLLOCATION TEST GAIN SCORES	370
TABLE A.28. A2 AND B1 PROGRESS VOCABULARY TEST SCORES (TIME 2)	370
TABLE A.29. A2 AND B1 PROGRESS COLLOCATION TEST SCORES (TIME 2)	370

List of Figures

FIGURE 2.1. THE ARCHITECTURE OF THE WORKING MEMORY MODEL (FROM BADDELEY, 2000: 418)	22
FIGURE 2.2. THE STRUCTURE OF THE PHONOLOGICAL LOOP (BADDELEY, 2003A: 193)	24
FIGURE 2.3. OUTLINE OF THE BURGESS AND HITCH (1999) STM MODEL	34
FIGURE 2.4. OUTLINE OF THE HARTLEY AND HOUGHTON (1996) STM MODEL.....	37
FIGURE 2.5. GUPTA AND MACWHINNEY (1997) STM MODEL.....	39
FIGURE 3.1. FREQUENCY PROFILE OF A TYPICAL LEARNER (MEARA, 1992: 4).....	74
FIGURE 3.2. FREQUENCY PROFILE OF GREEK LEARNERS OF EFL (MILTON, 2006C)	75
FIGURE 3.3. INDIVIDUAL FREQUENCY PROFILES OF 10 HIGHEST SCORING LEARNERS OF EFL IN MILTON (2006C)	76
FIGURE 3.4. CHECKLIST FORMAT IN THE X-LEX VOCABULARY TEST (FROM MILTON, 2009: 254).....	80
FIGURE 4.1. TERMS USED TO REFER TO VARIOUS ASPECTS OF FORMULAIC LANGUAGE (WRAY, 2002: 9)	110
FIGURE 5.1. PILOTING – BOXPLOT OF THE DISTRIBUTION OF L1 VOCABULARY TEST SCORES (N=14)	171
FIGURE 7.1. BOXPLOTS OF SERIAL NONWORD RECALL AND RECOGNITION SCORES IN A2 AND B1 GROUPS (TIME 1)	201
FIGURE 7.2. BOXPLOTS OF SERIAL NONWORD RECALL AND RECOGNITION SCORES IN A2 AND B1 GROUPS (TIME 2)	203
FIGURE 7.3. BOXPLOTS OF L1 VOCABULARY TEST SCORES IN A2 AND B1 GROUPS (TIME 1).....	204
FIGURE 7.4. BOXPLOTS OF EXPOSURE SCORES IN A2 AND B1 GROUPS	205
FIGURE 7.5. BOXPLOTS OF A2 VOCABULARY LEVELS TEST SCORES (TIME 1 AND TIME 2)	206
FIGURE 7.6. BOXPLOT OF GAINS IN A2 VOCABULARY LEVELS TEST SCORES	206
FIGURE 7.7. BOXPLOTS OF B1 VOCABULARY LEVELS TEST SCORES (TIME 1 AND TIME 2).....	207
FIGURE 7.8. BOXPLOT OF GAINS IN B1 VOCABULARY LEVELS TEST SCORES.....	208
FIGURE 7.9. BOXPLOTS OF A2 GITSAKI COLLOCATION TEST SCORES (TIME 1 AND TIME 2)	209
FIGURE 7.10. BOXPLOT OF GAINS IN A2 GITSAKI COLLOCATION TEST SCORES	209
FIGURE 7.11. BOXPLOTS OF B1 GITSAKI COLLOCATION TEST SCORES (TIME 1 AND TIME 2) AND GAINS IN B1 GITSAKI COLLOCATION TEST SCORES	210
FIGURE 7.12. BOXPLOTS OF A2 PROGRESS VOCABULARY TEST SCORES (TIME 2) AND B1 PROGRESS VOCABULARY TEST SCORES (TIME 2)	211
FIGURE 7.13. BOXPLOTS OF A2 PROGRESS COLLOCATION TEST SCORES (TIME 2) AND B1 PROGRESS COLLOCATION TEST SCORES (TIME 2)	212
FIGURE 7.14. CROSS-LAGGED PANEL CORRELATION PARADIGM (FROM KENNY, 1975: 888).....	247
FIGURE 7.15. CROSS-LAGGED CORRELATIONS BETWEEN SERIAL NONWORD RECALL AND VOCABULARY/ COLLOCATION SCORES IN A2 AND B1 GROUPS	249
FIGURE 7.16. CROSS-LAGGED CORRELATIONS BETWEEN SERIAL NONWORD RECOGNITION AND VOCABULARY/ COLLOCATION SCORES IN A2 AND B1 GROUPS.....	251
FIGURE 8.1. DEPICTION OF THE RELATIONSHIP BETWEEN SERIAL NONWORD RECALL AND L2 VOCABULARY/COLLOCATION KNOWLEDGE/LEARNING IN ADULTS AT A2 AND B1 LEVELS OF L2 PROFICIENCY	271
FIGURE A.1. PILOTING – NORMAL PROBABILITY PLOT OF SERIAL NONWORD RECALL SCORES (N=20)	363
FIGURE A.2. PILOTING – NORMAL PROBABILITY PLOT OF SERIAL NONWORD RECOGNITION SCORES (N=20)	363

FIGURE A.3. PILOTING – NORMAL PROBABILITY PLOT OF POLISH VOCABULARY TEST SCORES (N=13) 364

FIGURE A.4. PILOTING - BOXPLOT OF THE DISTRIBUTION OF POLISH VOCABULARY TEST SCORES (N=13) 364

FIGURE A.5. PILOTING – NORMAL PROBABILITY PLOT OF A2 GITSAKI COLLOCATION TEST SCORES (N=15) 365

FIGURE A.6. PILOTING – NORMAL PROBABILITY PLOT OF B1GITSAKI COLLOCATION TEST SCORES (N=15) 365

FIGURE A.7. PILOTING - BOXPLOT OF THE DISTRIBUTION OF A2 GITSAKI COLLOCATION TEST SCORES (N=15) 366

FIGURE A.8. PILOTING - BOXPLOT OF THE DISTRIBUTION OF B1GITSAKI COLLOCATION TEST SCORES (N=15) 366

FIGURE A.9. NORMAL PROBABILITY PLOT OF A2 SERIAL NONWORD RECALL SCORES AT TIME 1(N=30) 371

FIGURE A.10. NORMAL PROBABILITY PLOT OF A2 SERIAL NONWORD RECOGNITION SCORES AT TIME 1 (N=30)..... 371

FIGURE A.11. NORMAL PROBABILITY PLOT OF B1 SERIAL NONWORD RECALL SCORES AT TIME 1 (N=30)..... 372

FIGURE A.12. NORMAL PROBABILITY PLOT OF B1 SERIAL NONWORD RECOGNITION SCORES AT TIME 1 (N=30)..... 372

FIGURE A.16. NORMAL PROBABILITY PLOT OF A2 POLISH VOCABULARY TEST SCORES AT TIME 1 (N=30)..... 373

FIGURE A.17. NORMAL PROBABILITY PLOT OF B1 POLISH VOCABULARY TEST SCORES AT TIME 1 (N=30)..... 373

List of Equations

EQUATION 5.1. CORRECTING CORRELATION COEFFICIENTS FOR ATTENUATION148

EQUATION 5.2. CALCULATING GUTTMAN SPLIT-HALF RELIABILITY ESTIMATE.....149

EQUATION 5.3. CALCULATING THE SPEARMAN-BROWN CORRECTION FOR LENGTH.....149

EQUATION 5.4. CALCULATING ITEM DIFFICULTY FOR SMALL GROUPS OF SUBJECTS154

EQUATION 7.1. TESTING THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN CROSS-LAGGED CORRELATIONS.....247

Chapter 1: Introduction

Although the phonological loop forms only a relatively small part of working memory it is probably the most distinctive aspect of the Baddeley and Hitch (1974) multicomponent model, and the one that has been investigated in most detail. After more than 30 years it still generates controversy, and importantly, new and interesting data. (Baddeley, 2007: 15)

Phonological Short-term Memory (PSTM, also referred to as the phonological loop) is a component of Baddeley's fractionated Working Memory model (initially developed in 1974 by Baddeley & Hitch). The present dissertation explores the link between L1-based (Polish) PSTM measures and L2 (English) vocabulary and collocation learning in a longitudinal classroom-based study involving young adult native speakers of Polish resident in an English-speaking country (Ireland). As part of this study we investigate six research questions, three of which examine the link between PSTM and the development of L2 vocabulary knowledge in adult learners (Questions 1, 2 and 3 discussed below), while the remaining three examine the link between PSTM and the development of L2 collocational knowledge in adult learners (Questions 4, 5 and 6).

The PSTM component of the Working Memory model (1974) is theorised to consist of two subcomponents: the phonological store and articulatory rehearsal mechanism. The phonological store is seen as responsible for holding memory traces for up to a few seconds before they fade, and the articulatory rehearsal mechanism is believed to reactivate the traces in order to prevent them from decaying. To compare the extent to which these two components are involved in adult L2 learning in this study, individual differences in PSTM are tapped by an L1-based Serial Nonword¹

¹ Nonwords are phonotactically legal made-up words that do not exist in the language they are based upon. They include only those sounds or series of sounds that exist in that language. "Raspuko" is an example of a Polish nonword, while "darrock" is an example of an English nonword.

Recall task and an L1-based Serial Nonword Recognition task². The Serial Nonword Recall is believed to tap both subcomponents, i.e. the phonological store and the articulatory rehearsal mechanism, while the Serial Nonword Recognition is believed to tap the phonological store only (Baddeley, 2003a).

In the current study L2 vocabulary and collocational learning is analysed within two acquisitional contexts – outside the classroom and within the classroom setting. To distinguish between these two contexts, we measured the learning of high-frequency L2 vocabulary other than that introduced during the TCD language course that our subjects were enrolled on³ and the learning of course-related L2 vocabulary. We also measured the learning of high- and low-frequency collocations other than those introduced during the TCD language course and the learning of course-related L2 collocations.

The study brings together a set of crucial features which have to date not been assembled in a single study in PSTM research with respect to SLA in adult learners. The features in question include:

- a) an investigation of PSTM effects on L2 vocabulary and collocation knowledge/learning in adults at different L2 proficiency levels (beginners and pre-intermediate),
- b) the use of a recall-based nonword measure of PSTM alongside a recognition-based nonword measure of PSTM in adults to assess the involvement of the phonological store and the articulatory rehearsal mechanism in L2 vocabulary and collocation knowledge/learning,
- c) the use of a frequency-based L2 vocabulary size measure to tap L2 vocabulary learning outside the classroom, and the use of an end-of-course L2 vocabulary measure to tap L2 vocabulary learning in class,
- d) the use of a general L2 collocation measure to tap L2 collocation learning outside the classroom, and the use of end-of-course L2 collocation measure to tap L2 collocation learning in class,

² Please note that for ease of reading the names of PSTM tasks are capitalised throughout. Please consult Chapter 2, Section 2.6, for a more detailed explanation as to why these two task formats were selected for the present project.

³ An L2 (English) course was taught at Trinity College Dublin between January and July 2008. The course will henceforth be referred to as the TCD language course.

e) the analysis of the causal direction between PSTM and L2 vocabulary/collocation measures in a longitudinal classroom-based project within a cross-lagged correlation paradigm,

and f) repeated measurements of PSTM capacity using L1-based nonword tasks (recall and recognition measures) over time in order to establish whether L1-based PSTM scores change over time in adults.

1.1. Research questions

This study examines six questions. These are set out below, along with a brief account in each case of the relevant research context, as well as an indication of the expected contribution of the present study.

***Question 1:** Does the relationship between PSTM and L2 vocabulary learning (in class and outside the classroom) decrease with an increased level of L2 proficiency in adults, as has been shown to happen in the case of child L1 learning?*

Laboratory-based studies on adults indicate that there is a strong link between PSTM (mostly tapped by Nonword Repetition) and L2 vocabulary learning at the onset of language learning (Gupta, 2003; Papagno & Vallar, 1992; Speciale et al., 2004: Experiment 1; see also the older adult part of the Service & Craik study, 1993). Two existing classroom-based longitudinal studies (O'Brien et al., 2006; Speciale, Ellis & Bywater, 2004: Experiment 2) present a mixed picture in respect of the same relationship between PSTM and adult L2 vocabulary learning in the classroom context. O'Brien et al. demonstrate that L1-based Serial Nonword Recognition scores at Time 1 correlate with L2 free productive knowledge of vocabulary at Time 2 in lower ability learners, but not higher ability learners (the exact proficiency level is not specified). Speciale et al., on the other hand, detect no significant link between their Nonword Repetition task (based neither on L1 nor L2) and receptive L2 vocabulary knowledge in adults at the A1 level of L2 proficiency. The difference between the results obtained by O'Brien et al. and those obtained by Speciale et al. may have been brought about by the nature of the different L2 vocabulary measures used in the two studies. O'Brien et al. measured free productive knowledge of L2 vocabulary in spoken output, while Speciale et al. measured the students' ability to arrange some high frequency words into 10 pairs that could be reasonably contrasted. As only certain words can be

contrasted in terms of their meaning, it is not unlikely that the Speciale et al. measure of receptive L2 vocabulary was too restricted in its content to reflect the amount of progress their students made over a 7-week period of language instruction.

Neither O'Brien et al. (2006) nor Speciale et al. (2004: Experiment 2) deal with the link between PSTM and different levels of L2 proficiency. If the pattern observed by O'Brien et al. among their high and low ability adult subjects is replicable in adult A2 and B1 learners, then we would expect a stronger relationship between L1-based PSTM measures and L2 vocabulary learning at the A2 level and a weaker (or no) relationship at the B1 level. If, however, the pattern observed by Speciale et al. in their adult A1 learners is replicable, then we would expect no relationship between L1-based PSTM measures and L2 vocabulary learning at the A2 level or at the B1 level.

This study explores the link between PSTM and L2 vocabulary knowledge/learning using a PSTM measure that involves articulation (like in the Speciale et al. study, Experiment 2) and a PSTM measure that does not involve articulation (like in the O'Brien et al. study). It also includes L2 vocabulary measures that are of a more general nature (like in the O'Brien et al. study) and L2 vocabulary measures that test the content of the language course only (like in the Speciale et al. study, Experiment 2). It explores the issue of the relationship between PSTM and L2 vocabulary knowledge/learning at different L2 proficiency levels, as the issue has not yet been addressed in respect of adult L2 learners.

Question 2: *Which component of PSTM is more closely linked to L2 vocabulary learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?*

A number of tasks have been used in PSTM research to measure individual differences in PSTM capacity, including Nonword Repetition, Serial Nonword Recall and Serial Nonword Recognition. Within the Working Memory model the Serial Nonword Recognition task is believed to mainly involve the nonarticulatory phonological store, while the Nonword Repetition and the Serial Nonword Recall tasks are assumed to involve both the store and the articulatory rehearsal component (Baddeley, 2003a). The nonword measures that involve articulation of presented

nonwords, such as Nonword Repetition or Serial Nonword Recall, are arguably the most frequently used memory tasks in PSTM research (see e.g. Gathercole, 2006a, for an overview of studies using Nonword Repetition). A nonword measure that does not involve articulation of presented nonwords, namely Serial Nonword Recognition, has been used in some L1 studies involving children (Gathercole & Pickering, 1999; Gathercole et al., 2001; Gathercole et al., 1999b; Thorn, Gathercole & Frankish, 2002) and in one L2 longitudinal study involving adult learners (O'Brien et al., 2006).

In a longitudinal study by Gathercole et al. (1999b) of young children (aged 4 at the onset of the study) the Serial Nonword Recognition task (Time 1) was shown to correlate more strongly with L1 vocabulary knowledge (Time 2) as compared to the correlation between the Nonword Repetition (Time 1) and L1 vocabulary knowledge (Time 2). Gathercole et al. interpret the fact that both Nonword Repetition and Serial Nonword Recognition correlated significantly with their L1 vocabulary measure as a support for the claim that it is the phonological store rather than the articulatory component that constrains child L1 word learning (Gathercole et al., 1999b: 65).

It is not clear whether or not Serial Nonword Recall and Serial Nonword Recognition make the same or different predictions in relation to L2 learning in adult learners, as no longitudinal study has yet compared the strength of the link between recall- and recognition-based measures and adult L2 vocabulary learning in a classroom setting. If the patterns of correlations obtained in young children (Gathercole et al., 1999b) are also observed in adults, then we would expect our Serial Nonword Recognition measure to be more closely linked with L2 vocabulary measures than our Serial Nonword Recall measure. As Gupta (2003) points out, however, it might be plausibly expected that the patterns of relationships observed between PSTM measures and word learning in children do not resemble those in normal adults simply because of differences between the two populations in the use of strategies in recall- and recognition-based PSTM tasks. The fact that adults are adept at using rehearsal strategies, while young children are not (Gathercole & Hitch, 1993; Pickering, 2006: 245; Pressley & Van Meter, 1992) may have an impact on the strength of the link between the rehearsal component of PSTM (tapped by Serial Nonword Recall) and L2 vocabulary learning in adults. An experimental study by Martin (2009: 87) of adult artificial language learning provides preliminary evidence that the link between recall-

and recognition-measures in adults may be different from that observed in children. The study, for example, indicates that the correlation between some lexical measures and Serial Nonword Recognition is weaker than with Nonword Repetition.

Question 3: *In the case of adults (as opposed to children who were the focus of previous research), what is the causal direction of the relationship between PSTM and the learning of high-frequency L2 vocabulary (outside the classroom) at the A2 and B1 levels of L2 proficiency? Is it the L1-based nonword scores at Time 1 that contribute to L2 Vocabulary Levels Test scores at Time 2, or is it L2 Vocabulary Levels Test scores at Time 1 that contribute to the L1-based nonword scores at Time 2?*

Many studies linking L1-based PSTM measures with L1/L2 development are correlational in nature. Since correlation does not imply causation, the direction of causation in this link is still open to debate. Even though PSTM is generally regarded as a strong *predictor* of word learning (e.g. Baddeley, 2003b; Papagno, Valentine & Baddeley, 1991; Service, 1992), in principle it is plausible that good PSTM functioning supports vocabulary development or that the reverse holds true. One of the arguments put forward in favour of the view that individual differences in linguistic knowledge drive PSTM performance is that PSTM measures are known to be influenced by long-term memory. The effects of ‘lexicality’ (better recall for words than nonwords) and ‘wordlikeness’ (better recall for nonwords resembling real words) indicate that at a recall stage serial memory for items is affected by top-down redintegrative processes⁴ drawing on long-term lexical knowledge (Gathercole et al., 2001). Bearing the above in mind, individuals with lower vocabulary skills might be expected to obtain lower PSTM scores as compared to those with better developed lexicons, while individuals with a richer knowledge of languages (L1 and others) may be expected to have more efficient PSTM.

One frequently cited study that addressed the issue of the direction of causation in the link between PSTM and vocabulary learning is a longitudinal study by Gathercole et al. (1992) on young children learning their L1. A cohort of 80 4-year olds from a middle class background was subjected to PSTM and receptive vocabulary testing. The tests were re-administered when the children reached the ages of 5, 6 and

⁴ Redintegration is a process whereby long-term memory representations influence short-term recall. In other words, phonological representations available in long-term memory are used to repair partially decayed traces of words (see Chapter 2, Section 2.3.2).

8. To explore the directionality of the association Gathercole and colleagues applied a cross-lagged partial correlational analysis, which involved calculating and comparing the strength of the correlations between early PSTM and later receptive vocabulary, and between early vocabulary and later PSTM scores. According to the assumptions of a cross-lagged correlational analysis, the correlation should be stronger in the causal than in the noncausal direction (e.g. Campbell, 1988; Campbell & Stanley, 1963; Crano, Kenny & Campbell, 1972; Crano & Mellon, 1978). The pattern obtained for the link between PSTM and vocabulary at ages 4 and 5 suggests that PSTM capacity drives vocabulary learning rather than vice versa. However, this pattern shifted for age range 6 – 8 suggesting that expanding vocabulary knowledge might drive PSTM performance at higher levels of familiarity with the language. French (2003 & 2004) also used a cross-lagged partial correlational analysis to investigate the directionality of the link between PSTM (Arabic Nonword Repetition) and L2 learning in 54 French children at a beginning level of L2 proficiency (presumably A2) enrolled in a 5-month intensive English course in Quebec's Saguenay Region (mean age = 11.2 years). Partial cross-lagged correlations revealed that PSTM played a causal role in the children's L2 development over time (including the development of L2 vocabulary), which is in line with the Gathercole et al. study (1992) in relation to the relationship between PSTM and L1 vocabulary learning at ages 4 and 5. In this thesis we use a cross-lagged correlational analysis to explore the causal direction of the association between PSTM and vocabulary learning, but this time in the case of adults at elementary (A2) and pre-intermediate (B1) levels.

Question 4: *In the case of adults, is there a relationship between PSTM and the learning of L2 collocations (in class and outside the classroom) at the A2 and B1 levels of L2 proficiency? If so, does the relationship between PSTM and L2 collocation learning decrease with an increased level of L2 proficiency?*

To the best of our knowledge, there are no longitudinal studies on children or adults that deal with the link between PSTM and L1/L2 collocational knowledge. A laboratory-based study by Ellis and Sinclair (1996) looked at how interfering with PSTM by means of articulatory suppression influenced the efficiency of learning new L2 utterances. The subjects who were prevented from repeating novel L2 expressions (the articulatory suppression condition) were significantly worse than those who were

instructed to remain silent (the silent condition) and those who were requested to repeat novel L2 expressions out loud (the repetition condition). The above results led Ellis and Sinclair to conclude that PSTM is implicated in the acquisition of L2 phrases in adults (1996: 245, see also Ellis, 1996 and Ellis & Schmidt, 1997). Ellis and Sinclair also suggest that the involvement of PSTM in L2 collocation learning resembles that for L2 vocabulary learning, at least at the very initial stage of L2 learning. Further support for the link between individual differences in PSTM capacity and the learning of L2 sequences in adults at the initial stages of L2 learning comes from the Martin (2009: 90) study, which looked at the link between PSTM (Nonword Repetition and Serial Nonword Recognition) and the learning of grammar in an artificial foreign language.

Some longitudinal studies on the role of PSTM in L1/L2 vocabulary learning (Gathercole et al., 1992; O'Brien et al., 2006) indicate that PSTM involvement in vocabulary learning decreases as L1/L2 proficiency increases. Gathercole (2006a: 515) points out that, with respect to the learning of individual words, more advanced L2 learners become less reliant on PSTM and more adept at using their mental lexicons to access phonological representations of close neighbours of the target item. The Ellis and Sinclair (1996) lab-based study deals with an initial stage of L2 learning, and it is not clear how their findings might extend to other levels of L2 proficiency. A greater learning effort can be expected in the case of multi-word phrases when compared to the learning of single words (Verstraten, 1992: 28), and therefore it is plausible that PSTM may still play a role in the learning of collocations at post-elementary levels of proficiency. Indirect support for this view comes from the O'Brien et al. (2006) study, in which a significant link between L1-based Serial Nonword Recognition (Time 1) with L2 subordinate clause use (Time 2) was detected in high and low ability students. On the basis of the above O'Brien et al. conclude that at initial stages of language learning PSTM is mainly deployed for vocabulary learning, but that at later stages when vocabulary access is more automatic PSTM is redeployed for learning more complex structures (2006: 399).

Question 5: *Which component of PSTM is more closely linked to L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?*

There is no research so far linking the Serial Nonword Recall task and the Serial Nonword Recognition task with L2 collocation learning in children and adults. Considering the fact that O'Brien et al. detected a significant correlation between their L1-based Serial Nonword Recognition task (Time 1) and L2 subordinate clause use (Time 2) in both low ability and high ability adult L2 learners, we could expect to obtain a similar pattern in relation to our L1-based Serial Nonword Recognition measure and L2 collocation measures (Time 2) at both the A2 and B1 levels of L2 proficiency. Also, if the finding by Martin (2009: 90) that Nonword Repetition, as compared to Serial Nonword Recognition, is more strongly linked with grammatical measures in adult artificial foreign language learning is replicable in a classroom setting, we could expect a stronger link between our Serial Nonword Recall measure and the learning of L2 collocations than between our Serial Nonword Recognition measure and the learning of L2 collocations.

If it is the nonarticulatory storage component of PSTM that is more closely associated with L2 collocation learning in adults, then we would expect L2 collocation scores to be significantly linked with both the Serial Nonword Recall and Recognition tasks (irrespective of their spoken recall requirement). If it is the articulatory component of PSTM that is more closely associated with L2 collocation learning in adults, we would expect L2 collocation scores to be significantly linked with the Serial Nonword Recall, but not with the Serial Nonword Recognition. If both of our PSTM measures correlate with L2 collocation measures, we shall also look at the patterns of obtained correlations in order to establish whether or not there are any significant differences in the strength of correlations among our PSTM measures and L2 collocation measures.

Question 6: *What is the causal direction of the relationship between PSTM and L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency? Is it PSTM scores at Time 1 that contribute to Gitsaki's Collocation Test scores at Time 2, or is it Gitsaki's Collocation Test scores at Time 1 that contribute to the PSTM scores at Time 2?*

As mentioned before, a longitudinal study by Gathercole et al. (1992) looked at the causal direction of the relationship between PSTM (L1-based Nonword Repetition task) and L1 vocabulary learning in children. Their cross-lagged partial correlational

analysis indicates that at ages 4 and 5 PSTM drives L1 vocabulary learning, while at ages 6 and 8 it is L1 vocabulary knowledge that contributes to PSTM scores. This suggests that at higher levels of familiarity with L1, expanding L1 vocabulary knowledge (Time 1) may drive PSTM performance (Time 2). If the same pattern applies to L2 collocation learning in adults, at the A2 level of L2 proficiency we would expect stronger correlations between PSTM (Time 1) and A2 Gitsaki's Collocation Test (Time 2) than between A2 Gitsaki's Collocation Test (Time 1) and PSTM (Time 2), and at the B1 level of proficiency we would expect the pattern to be reversed.

1.2. Operational definitions

The following operational definitions are assumed throughout the thesis, namely Phonological Short-term Memory, L2 vocabulary learning, L2 collocational learning, L1 vocabulary knowledge and exposure to L2 outside the classroom.

1.2.1. Phonological Short-term Memory (PSTM)

For the purpose of this study two operational definitions of PSTM capacity were formulated – one involving articulation of lists of nonwords⁵ and the other involving passive recognition sets of nonword lists. The two operational definitions of PSTM adopted here are thus as follows:

1. Serial Nonword Recall performance (the ability to retain and repeat L1-based nonword lists of varying lengths immediately after the presentation of each list with the correct nonword order maintained) – at Time 1 and Time 2⁶,
2. Serial Nonword Recognition performance (the ability to retain sets of two L1-based nonword list of varying lengths and to judge, immediately after the presentation of a given set, whether the nonwords within each set are presented in the same order) – at Time 1 and Time 2.

⁵ Nonwords in the study were novel combinations of L1 (Polish) syllables.

⁶ Time 1 is used to refer to the commencement of the language course, while Time 2 denotes the termination of the course.

1.2.2. L2 vocabulary knowledge and learning

The second set of concepts to be defined operationally are the notions of vocabulary knowledge and vocabulary learning. In many frameworks (e.g. Nation, 1990; Qian, 2002) there is a consensus that vocabulary knowledge comprises at least two dimensions, that is vocabulary breadth/size and vocabulary depth. Vocabulary breadth refers simply to the number of words that the learners are familiar with (even if the understanding of the meaning is only superficial), while depth of vocabulary refers to the learners' knowledge of various aspects of words under investigation, such as e.g. word associations (see Chapter 3, Section 3.2 for more detail). This study focuses on gains in high-frequency L2 vocabulary breadth (i.e. selected L2 words from frequency lists containing the first, second and third 1,000 most frequent L2 words) and the breadth of L2 vocabulary learnt in class (i.e. selected L2 words that were introduced in class and recycled at least twice).

The operational definition of **L2 vocabulary knowledge** used in the current study is as follows:

L2 vocabulary knowledge is defined as that which has been gained outside the classroom (measured in relation to L2 vocabulary that has not been explicitly taught during the TCD language course):

- knowledge of the receptive and controlled productive general vocabulary breadth defined in relation to the English word frequency lists used in the creation of Vocabulary Levels Test (Nation, 2001b; Schmitt, Schmitt & Clapham, 2001) – at Time 1 and Time 2

The operational definitions of **L2 vocabulary learning** used in the current study include the following:

1. L2 vocabulary learning which has taken place outside the classroom:

- gains in the receptive and controlled productive general vocabulary breadth defined in relation to the English word frequency lists used in the creation of Vocabulary Levels Test (Nation, 2001b; Schmitt et al., 2001) – between Time 1 and Time 2

2. L2 vocabulary learning which has taken place in class:

- the receptive and controlled productive knowledge of vocabulary that was introduced in class and recycled at least twice (breadth test format) – and which had been developed by Time 2

1.2.3. L2 collocation knowledge and learning

The third set of concepts to be defined operationally are the notions of collocations and collocation learning. As outlined in Chapter 3, the major conceptual approaches to collocations include the frequency-based view and the phraseological view of collocations. In the frequency-based tradition collocations are perceived as words co-occurring within a certain distance from one another “with a probability greater than chance” (Halliday, 1966) regardless of their idiomaticity and their structural status (Biber et al., 1999: 990). In the phraseological tradition, on the other hand, attention is paid to collocation types rather than their frequency of occurrence (Benson et al., 1986; Benson et al., 1997), and to various degrees of fixedness and opaqueness (Howarth, 1998). The approach adopted in this thesis is embedded in the frequency-based tradition. The frequency of occurrence of target collocations was calculated on the basis of the British National Corpus using the BNCweb interface (Hoffmann & Evert, 2007). The target collocations included a number of grammatical and lexical types of collocations (BBI typology by Benson et al., 1986).

Our approach towards collocations overlaps with that of Gyllstad, who defines them as:

(...) two syntagmatically related and frequently co-occurring orthographic words, either adjacent or separated by a specified distance (...) where the meaning is evoked by the combination as a whole, sometimes requiring additional lexical elements for grammatical well-formedness and usage convention. (Gyllstad, 2009: 155)

For two orthographic words to be classified as a collocation in this study they had to fulfil all of the following criteria:

1. be two-word units in which the co-occurring item appears within 3 words to the left or right of the node,
2. be listed in at least one of the following dictionaries: “*Oxford Collocations: Dictionary for Students of English*” (2002), “*The BBI Combinatory Dictionary of English*” (Benson et al., 1986), “*The BBI Dictionary of English Word Combinations*” (Benson et al., 1997) or “*Oxford Advanced Learner’s Dictionary*” (Hornby, 2005),
3. be present in the BNC corpus (no minimum threshold of frequency of occurrence was set).

Selected two-word units that fulfilled the above requirements were classified as a) low- and high-frequency collocations in accordance with their BNC frequencies of occurrence and as b) stronger and weaker collocations according to their Mutual Information cubed (MI3) scores also calculated on the basis of BNC data.

L2 collocation knowledge in the current study refers to the following:

L2 collocation knowledge gained outside the classroom (measured in relation to L2 collocations that have not been explicitly taught during the TCD language course):

- controlled productive knowledge of L2 collocations measured by the A2 and B1 Gitsaki’s Collocation Tests – at Time 1 and Time 2.

L2 collocational learning in the current study refers to the following:

1. L2 collocation learning which has taken place outside the classroom:
 - gains in the productive knowledge of L2 collocations measured by the A2 and B1 Gitsaki’s Collocation Tests (1999)⁷ – between Time 1 and Time 2,

⁷ The target collocations on Gitsaki’s collocation tests come from A2 and B1 English language teaching materials other than those used during the TCD language course that our subjects were enrolled on.

2. L2 collocation learning which has taken place in class:

- receptive and controlled productive knowledge of collocations that were introduced in class and recycled at least twice – and which had been developed by Time 2.

1.2.4. L1 vocabulary knowledge

Productive knowledge of L1 (Polish) low-frequency vocabulary was included in our project as a potentially confounding variable. L1 long-term knowledge is assumed to be linked to the process of creation of L2 phonological representations in adult L2 learners at the onset of L2 learning, provided that the perceived cross-language phonetic distance between is not considerable (Flege & MacKay, 2004). It is plausible that in languages which are close to one another phonetically, such as for example Polish and Czech, the activation of L1 long-term knowledge at initial levels of proficiency may be sufficient to compensate for individual differences in PSTM functioning. In the case of Polish and English, however, it is not clear to what extent knowledge of Polish phonological regularities would interact with the learning of novel English forms, and therefore we feel it is essential that a Polish vocabulary measure be incorporated in the experimental design.

L1 vocabulary knowledge was in this study defined as L1 low-frequency vocabulary breadth (at Time 1). For L1 nouns, verbs and adjectives to be defined as low-frequency they had to have a frequency of less than 600 of occurrences in the IPI PAN corpus (2nd edition, 2006), a 250 million word corpus of written Polish language.

1.2.5. Exposure to L2 outside the classroom

Another potentially confounding variable identified at the design stage of our project was the amount of exposure to L2 outside the classroom. Our subjects were resident in Dublin (Ireland) throughout the duration of the TCD language course and were therefore exposed to some amount of L2 outside the classroom. Since Poles were the most dominant migrant group in Ireland in the two years preceding 2008 (see Chapter 6, Section 6.2.1), we could not assume that the patterns of exposure to L2 would be similar for all of our subjects. A considerable number of the Polish migrants

that we had interviewed in 2006 and 2007 under an umbrella of the larger Polish Diaspora Project reported that they could go about their daily routines interacting mainly in L1. They pointed to the fact that a number of services were available in Polish (Polish shops, Polish schools, Catholic masses in Polish, Polish legal advice, Polish film festivals, etc.). Some of them admitted to socialising mainly with other members of the Polish community in Ireland owing to a common linguistic and cultural background. Among the Poles we had interviewed prior to 2008, there were also a number of individuals who reported interacting through L2 mainly and using hardly any L1 on a daily basis.

In the current study exposure (to L2 outside the classroom) was defined as:

- interacting in L2 with L2 speakers – face-to-face, via the phone or on the Internet (chat rooms, e-mails),
- watching TV in L2, reading books, newspapers, etc. in L2, listening to radio programmes in L2.

Exposure was measured in terms of hours per day between Time 1 and Time 2. The subjects were asked to keep a diary for six weeks (one selected week of each month during the TCD language course), in which they were requested to note the number of hours of exposure to L2 outside the classroom per day⁸. Our subjects were instructed not to include any TCD language course-related activities (such as homework) in their estimations of exposure.

⁸ Please note that the Exposure measure reported in our 2009 article in *EUROSLA Yearbook 9* was a different measure. That measure took account of self-reported estimation of the ratio between the use of L1 (Polish) and L2 (English) in the subjects' daily lives, and was calculated on the basis of the answers to Question 25 from our background questionnaire (Appendix B). The Exposure measure reported here is meant to reflect the average number of hours our subjects were exposed to L2 outside the classroom between Time 1 and Time 2.

Chapter 2: Phonological Short-term Memory (PSTM) and Adult Second Language Learning

2.1. Introduction

Over the last 30 years the Baddeley and Hitch 1974 Working Memory model has consistently been at the centre of SLA studies and has proved to be of both theoretical and practical relevance. It has been depicted as “a major accomplishment of the modern era of cognitive psychology, (...) one of the first information processing models to make an explicit attempt to accommodate both traditional behavioural data and the results of neuropsychological case studies in an integrated theoretical framework” (Buchsbaum & D'Esposito, 2008: 762). The phonological loop component of the Working Memory model (Phonological Short-term Memory, PSTM), assumed to consist of a phonological store and an articulatory rehearsal mechanism, is the central focus of the thesis.

Section 2.2 focuses on how the development of the concept of working memory and short-term memory up to the 1970s contributed to the formulation of the Baddeley and Hitch Working Memory model. Section 2.3 outlines the architecture of the Working Memory model, focusing in particular on the phonological loop component (PSTM). It also highlights the strengths and weaknesses of the model in the light of the current evidence for the existence of PSTM. Since the Baddeley and Hitch Working Memory model does not directly address the mechanisms responsible for the interaction between PSTM and LTM, Section 2.4 presents a number of models that simulate PSTM computationally. Section 2.5 contains a brief overview of research on the association between PSTM and L1/L2 vocabulary learning. Section 2.6 describes Serial Nonword Recall and Serial Nonword Recognition, the two measures employed in our research project, and their theoretical relevance within the Working Memory model. Since our PSTM tasks contain L1-based nonwords, the final section (2.7) deals with the issue of the impact of language of testing on PSTM scores.

2.2. The evolution of the concept of Working Memory and Short-term Memory

The internal structure of memory and its impact on how incoming sensory information is processed and retained has been a matter of considerable interest among cognitive psychologists and psycholinguists. The majority of cognitive psychologists subscribe to the view that memory can be divided into two inter-dependent components, that is a limited capacity short-term memory (STM) store (or working memory, WM) and a greater capacity, more durable long-term memory (LTM) store. To contextualize the notion of Phonological Short-term Memory (PSTM), the evolution of STM, WM and LTM is briefly presented in this section.

One of the earliest references to the concept of short-term memory (STM) comes from the writings of a British philosopher John Locke (1690), who distinguished between temporary memory for “the idea that is brought into it” and a more permanent “storehouse of ideas” (ibid.: Book II, Chapter X, paras. 1-2). The same type of dichotomy appeared again in the work of an American philosopher and psychologist, William James (1890/1931), who coined the term primary memory in relation to what he termed as the “feeling of the specious present” (ibid.: 647), and secondary memory in relation to “knowledge of a former state of mind after it has already once dropped from consciousness” (ibid.: 648). Primary memory was regarded as being a more faithful recording of memory traces, as it reflected an event that was part of the psychological present. Secondary memory, on the other hand, reflected an event belonging to the psychological past and, hence, was more likely to abound with distortions and omissions.

James’ assumption that STM and LTM are separate systems was reinforced by subsequent computer models, in which limited capacity STM was typically included as a route into LTM (Murdock, 1974). One of such models, for example, is Broadbent’s tripartite information-processing model (1958). It comprised three main memory storages – two temporary storage systems and a larger capacity LTM system – as well as a rehearsal loop mechanism. The two temporary storage systems included the S-system, a preattentive sensory store, and the P-system, a limited-capacity store that received filtered information from the S-system. The rehearsal loop mechanism was responsible for maintaining information in STM via constant data transfer between the

temporary storage systems. While the P-system corresponds roughly to a STM store in subsequent memory models, such as e.g. the modal model of Atkinson and Shiffrin (1968), the concept of rehearsal mechanism can be seen as an underlying foundation for Baddeley and Hitch's articulatory rehearsal mechanism within their Working Memory Model (1974 and later versions).

Primary and secondary memory as well as the nature of the rehearsal mechanism were further refined in two very influential papers, one by Waugh and Norman (1965) and the other by Atkinson and Shiffrin (1968). Waugh and Norman took the line that that rehearsal was essential not only in preventing the displacement of a stimulus within the primary store by other more recent stimuli, but also in transferring the stimulus from a limited capacity primary memory store to a more permanent secondary store. Unlike James, Waugh and Norman viewed primary memory in terms of the number of events that could be stored in STM rather than the period of time these events would extend over.

In the Atkinson and Shiffrin (1968) modal model, that comprised of sensory registers, STM and LTM, the STM store was believed to control the flow of information in and out of LTM through control processes such as rehearsal, coding and retrieval. The modal model resembled the Waugh and Norman model in that it posited that memory items, represented in STM in a phonological form, can be maintained through rehearsal processes or lost due to decay or interference. In contrast to the Waugh and Norman model, however, STM was believed to have not only storage but also processing functions. Atkinson and Shiffrin suggested that due to STM capacity limitations a trade-off between storage and processing can be expected to occur. The authors also briefly introduced the term *working memory* (WM), and they explicitly assumed that STM functioned as working memory. However, there was little empirical support for that assumption at the time. The question of whether or not STM functions as WM was later directly tested by Baddeley and Hitch (Section 2.3), which led to the development of their 1974 Working Memory Model.

Despite providing a good account of a considerable amount of empirical data available at the time, the modal model ran into a few problems and its detailed application has been criticised (Baddeley, 1978: 149; Baddeley, 2007: 4; Baddeley & Hitch, 2007: 2). One of the problems relates to Atkinson and Shiffrin's assumptions

about the impact of the interaction between STM and LTM on the learning process. Atkinson and Shiffrin posited that holding data in STM was sufficient for it to be transferred into LTM and that the longer the data was held in STM, the better long-term learning ensued. However, attempts to confirm whether the adequacy of an item's registration in LTM is indeed a positive function of its length of stay in STM have proved unsuccessful. As Craik and Watkins (1973) showed in their incidental learning task, neither increased length of an item's stay in STM nor the number of overt rehearsals the item received resulted in better recall.

The second problem with the modal model stemmed from neuropsychological evidence (e.g. Shallice & Warrington, 1970; Vallar & Shallice, 1990). Since access to LTM was represented as possible only through the STM store, the model's prediction was that patients with a massive deficit in STM should also encounter associated problems in long-term learning. Furthermore, since STM was assumed to act as working memory, patients with deficient STM would be expected to exhibit general information-processing deficits affecting their performance on many different cognitive tasks (Baddeley, 2007: 4; Tulving & Craik, 2005: 82-83). Those few relatively pure cases of selective STM deficits that were studied showed that patients with STM impairment can lead normal lives and perform effectively in their demanding jobs. The STM patients appeared to show no evidence of a general cognitive dysfunction, and no impairments of long-term learning were detected either (e.g. patient K.F in Shallice and Warrington's study, 1970). Such data were interpreted as an indication that models in which information must enter the STM store before reaching the LTM store (Atkinson & Shiffrin, 1968; Murdock, 1967; Waugh & Norman, 1965) may be inaccurate.

The proposed relationship within the modal model (Atkinson & Shiffrin, 1968) between the length of an item's stay in STM and the efficiency of its long-term learning prompted Craik and Lockhart (1972) to put forward their highly influential Levels-of-Processing Hypothesis. Contrary to Atkinson and Shiffrin, Craik and Lockhart argued that it is the depth and richness of encoding – rather than the length of an item's stay in STM – that determine the probability of long-term learning. Craik and Lockhart's theory predicted that deeper levels of processing, such as e.g. semantic judgement, result in better recall than shallow levels of processing, such as e.g. maintenance rehearsal. Rehearsal, however, has been shown to result in long-term learning and to

serve more than just a maintenance function (e.g. Baddeley, 1978), even though it was regarded by Craik and Lockhart as operating at a shallow level. This finding led Baddeley and Hitch to look for a better explanation of the link between rehearsal and long-term learning.

2.3. The Working Memory model of Baddeley and Hitch

Baddeley and Hitch (1974) built their multi-component Working Memory (WM) model on the basis of the data obtained from dual-task studies, that is studies of normal subjects performing two tasks concurrently. According to dual-task methodology, if two tasks compete for a single limited-capacity WM system, the performance on one or both will be substantially disrupted under demanding dual-task conditions as compared to single-task baseline levels. The rationale behind this assumption is that, if the limited-capacity memory system is unitary and one of the tasks in hand requires most of its available capacity, there should be little remaining capacity to cope with additional loads, resulting in a significant decrease in performance. Interestingly, a range of dual-task studies revealed that some memory tasks can be performed simultaneously by normal adults without much disruption, even when each task on its own places very heavy demands on the memory system (e.g. Baddeley et al., 1986 & 1991; Baddeley & Hitch, 1974; Della Sala et al., 1995; Duff & Logie, 2001; Greene, Hodges & Baddeley, 1995), while other tasks caused substantial mutual interference when performed concurrently (e.g. Cocchini et al., 2002; Farmer et al., 1986).

The relative lack of dual-task interference in certain task combinations has come to be viewed as evidence in favour of a multi-component memory system with a number of separate specialized cognitive functions responsible for online processing and temporary storage. In the experiments conducted by Baddeley et al. (1991; 1986) two types of task were used, namely digit span (the maximum number of digits that could be recalled in the original order straight after the presentation) and tracking (the ability to follow a moving patch of light on a computer screen) to investigate whether the experimental tasks in question tapped the resources of same cognitive component. Digit span was seen as reflecting the capacity for retention of serial verbal information, while tracking was seen as tapping working memory capacity for visuospatial information and motor control. Since in the concurrent performance condition there

was very little observable drop in the performance of either of the tasks, the suggestion was put forward that the cognitive components these two tasks drew upon – verbal and visuospatial processing and storage – were separate and could work more or less independently of one another.

More evidence in support of the hypothesis of separate verbal and visuospatial stores comes from studies that use articulatory suppression, an experimental technique involving concurrent articulation of irrelevant verbal stimuli in order to make articulatory resources unavailable for rehearsal of to-be-remembered verbal material. In line with the interpretation of dual-task data, digit span was shown to be severely impaired by articulatory suppression (Murray, 1967), but visuospatial memory was not affected (Cocchini et al., 2002; Smyth, Pearson & Pendleton, 1988), suggesting that verbal and visuospatial tasks tap stores that are independent of one another.

When Baddeley and Hitch launched their research on the relationship between STM and LTM, the research on STM was beginning to lose its popularity in favour of LTM theories such as e.g. levels of processing (Craik & Lockhart, 1972). In order to explain patterns observed within their own dataset, Baddeley and Hitch (1974) replaced the notion of a unitary STM system, such as that in the Atkinson and Shiffrin (1968) modal model or that in the Levels-of-Processing model, with a memory model that consisted of multiple interdependent components. The framework they proposed has become widely known as the Working Memory (WM) model or the fractionated WM model. It has also been referred to as the European WM model, as it has been particularly dominant within Europe. Baddeley and Hitch used the term *working memory* in relation to “a limited capacity temporary storage system that underpins complex human thought” (Baddeley, 2007: 6-7). They stressed the functional role of the proposed system, rather than just its storage function, and continued to explore the architecture of the model and the respective functions of its components. As a result, the 1974 WM model has been refined over the years (Baddeley, 1986 & 2000; see also Baddeley 1990, 1992, 2003, 2007; Baddeley & Logie, 1992), but it needs to be pointed out that irrespective of minor modifications the relative simplicity of the architecture of the WM model has been retained.

2.3.1. The Architecture of the Working Memory model

The WM model consists of a supervisory attentional system, called the central executive, and three subsidiary systems – namely the phonological loop (PSTM), the visuospatial sketchpad and the more recently incorporated episodic buffer (Baddeley, 2000; see Figure 2.1 below). The central executive, probably “the most important but the least understood component of working memory” (Baddeley, 2003b: 835), is currently treated as a purely attentional system. The original view was that it comprised both attentional and storage functions. The episodic buffer, an additional storage component, is theorised to be either a fractionation of the original central executive (Baddeley, 2007: 147) or the fourth component of the WM model (Baddeley, 2000). The two remaining components, visuospatial sketchpad and the phonological loop, are separated into processing and storage systems. The visuospatial sketchpad is responsible for visual and spatial information, while the phonological loop, which is the focus of this research project, deals with the manipulation and retention of verbal material.

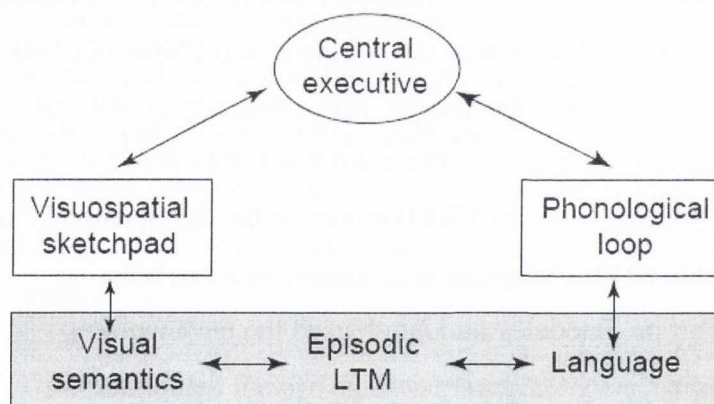


Figure 2.1. The architecture of the Working Memory model (from Baddeley, 2000: 418)

The phonological loop component of the WM model is assumed to comprise a phonological store, which can hold memory traces for up to a few seconds before they fade, and an articulatory rehearsal mechanism, which reactivates the traces and prevents them from decaying (e.g. Baddeley, 2003b). The phonological store has been compared to an ‘inner ear’ that stores speech-based information, while the articulatory rehearsal process has been likened to ‘inner voice’ that repeats verbal information on a loop to keep the trace active within the phonological store (Cohen, Eysenck & Le Voi,

1986). Some changes in terminology have occurred since the model was first introduced, but the essence of the phonological loop has been retained. Before it came to be known as *the phonological loop* (Baddeley, 1990), it was referred to as *the phonemic buffer/loop* (Baddeley & Hitch, 1974) and *the articulatory loop* (Baddeley, 1986). The term *phonological short-term memory* (PSTM) is also commonly used to refer to the functions that the phonological loop is assumed to have (e.g. Bishop, 2006; Gathercole & Thorn, 1998; Kormos & Sáfár, 2008; Masoura & Gathercole, 1999; Papagno & Vallar, 1992).

The core assumptions of the phonological loop model include a) the phonological coding of verbal sequences, b) transfer of visual code into phonological information via rehearsal, c) information loss through time-based decay or interference, and d) memory trace retention via rehearsal. The very first assumption relates to the fact that the phonological store component of the phonological loop is thought to be phonologically based. This is supported by the fact that its efficiency decreases substantially when phonologically similar stimuli are retained in the store (phonological similarity effect; Baddeley, 1966; Conrad & Hull, 1964). In other words, the level of phonological similarity between memory items predicts serial recall accuracy; that is, lists of items that sound similar are more difficult to recall than lists of items that sound different (Mueller et al., 2003).

Another core assumption of the phonological loop model is that auditorily presented verbal information gets direct access to the store, where it is retained in its phonological form, but visually presented verbal information has to be transformed into a phonological code via the rehearsal mechanism before it can be temporarily retained in the store (Baddeley, Lewis & Vallar, 1984; Baddeley, Thomson & Buchanan, 1975; Murray, 1967; see Figure 2.2 below). In principle, such phonological information could be retained indefinitely in the store via continuous rehearsal. If not refreshed via rehearsal, phonological representations are subject to decay over time as well as interference from newer items (Baddeley, 2007: 61). The rehearsal process is believed to involve either overt or covert (subvocal) articulation (Baddeley, 2007: 7). During subvocal articulation the same articulatory gestures that are involved in overt articulation are activated, but are not executed, which is why when we engage in

subvocal rehearsal we are able to rehearse verbal information without having to move our lips or any other parts of the articulatory apparatus (Gathercole, 1992: 54).

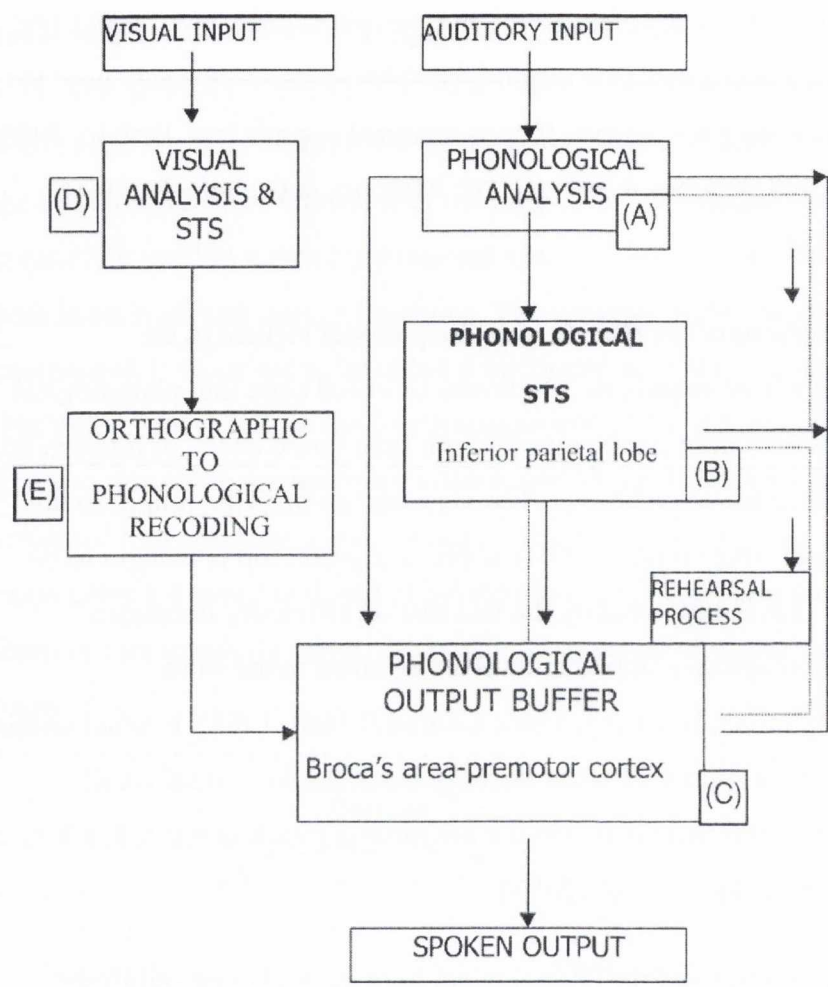


Figure 2.2. The structure of the phonological loop (Baddeley, 2003a: 193)

2.3.2. Current evidence for PSTM

Despite the fact that the phonological loop (PSTM) is a relatively small component of the WM model, it has remained theoretically and empirically productive over the last 30 years. This section focuses on the phenomena that give better insight into the functioning of the phonological loop, namely the phonological similarity effect, the word length effect, articulatory suppression, the irrelevant sound effect, the influence of LTM and redintegration.

Phonological similarity effect

A better understanding of the nature of the phonological loop has emerged from empirical data documenting the effect of phonological similarity in serial recall of verbal material. Typically, the recall of sequences of phonologically dissimilar words is higher than that of phonologically similar words (e.g. Baddeley, 1966; Baddeley, 1968; Conrad & Hull, 1964; Mueller et al., 2003; Schweickert, Guentert & Herberger, 1990). Conrad and Hull (1964) demonstrated that sequences of phonologically similar letters such as B D T G C P are recalled less accurately than those of phonologically dissimilar letters such as F K Y W R Q. Baddeley (1966) extended Conrad and Hull's finding to words, showing that the recall of sequences of phonologically similar words, such as e.g. *man cat cap map can*, can be as low as about 10 per cent, in contrast to dissimilar sequences, like e.g. *pit day cow pen sup*, which are recalled correctly in approximately 80 per cent of the occasions⁹.

The phonological similarity effect is interpreted as evidence supporting the phonological loop model's assumption that it is the phonological store that holds verbal information rather than any other kind of storage system e.g. visual or semantic. Within the phonological loop model there are at least two explanations of the observed phonological similarity effect. One possible explanation is that there is an interaction between phonological similarity and time of decay; that is, phonological codes held in STM decay more quickly when items to-be-remembered are phonologically similar (Posner & Konick, 1966). Another explanation could be that the rate at which the codes decay is roughly the same regardless of their phonological similarity, but partially degraded traces of phonologically similar items are more difficult to reconstruct or *redintegrate* at the time of recall than partially degraded traces of phonologically dissimilar items (Nairne, 1990; Service, 2007). According to the redintegration hypothesis (Hulme, Maughan & Brown, 1991) long-term phonological knowledge is used to compensate for loss of elements of longer words. Reconstructive mechanisms redintegrate partially degraded phonological traces on the basis of long-term knowledge about the constituent words. The process of redintegration is dealt with

⁹ For a more detailed account of the phonological similarity effect (including a discussion of some contradictory evidence) please consult e.g. Gupta, Lipinski and Aktunc (2005), Nimmo (2004), and Lewandowski and Farrell (2008).

in the section discussing the influence of long-term memory knowledge later in this chapter.

Word length effect and articulation rate

Another very robust effect of STM is the so-called word length effect, higher recall accuracy for lists of short words as compared to that for lists of long words. The first to report this effect was Baddeley et al. (1975), who compared memory spans for different lengths of to-be-remembered words. His study revealed that, even when frequency of occurrence and the level of abstractness were controlled for, there was a systematic decrease in memory span as the length of words increased. This effect was initially linked with articulatory durations of verbal stimuli (Baddeley et al., 1975). It was not the number of syllables per se that was considered critical, but rather the rate at which the item could be rehearsed to prevent it from decaying. The more time the word took to rehearse, the greater trace decay was believed to occur during rehearsal. Lengthy multisyllabic words were claimed to receive less rehearsal than shorter items, which lead to their more rapid decay. Consistent with this interpretation was the finding that speed of articulation of to-be-remembered items correlated with memory spans for the very same items (Hulme et al., 1984), and that a developmental increase in memory span among children showed a strong link with increases in articulation rates (Nicolson, 1981). Furthermore, among patients with impaired phonological loop (Belleville, Peretz & Arguin, 1992; Vallar & Baddeley, 1984) as well as among normal subjects engaged in irrelevant sound articulation (Baddeley et al., 1975) the word length effect was removed (see Baddeley, 2007 for a more detailed discussion).

Articulatory suppression

Disrupting the functioning of the phonological loop is another significant source of evidence regarding the nature of this component. There are a number of ways that have been used to interfere with phonological memory, including nonverbal interference (e.g. concurrent tapping), irrelevant sound condition (simultaneous presentation of irrelevant speech) and articulatory suppression (also referred to as concurrent articulation). Articulatory suppression, as briefly mentioned earlier, involves engaging in a concurrent verbal task while attempting to memorize the target memory items. Such concurrent articulation of irrelevant verbal material is assumed to

constitute a demanding secondary task that competes for the limited capacity of the phonological loop, and – unlike concurrent tapping, which is often used as a control condition – has a significant disruptive effect on STM measures (Levy, 1971; Murray, 1968; Peterson & Johnson, 1971).

Empirical findings from studies employing articulatory suppression contrast the phonological similarity effect (assumed to reside within the phonological store) and the word length effect (believed to reflect the process of rehearsal) in two modes of presentation, namely auditory and visual. With visually presented material articulatory suppression removes both the phonological similarity effect and the word length effect. However, when items are presented auditorily one of the two effects – the phonological similarity effect – is retained regardless of articulatory suppression. In contrast to visual presentation, the fact that the phonological similarity effect remains in the auditory mode is accounted for by the hypothesis that verbal material is registered within the phonological store directly and does not require subvocal rehearsal (Baddeley et al., 1984; Baddeley et al., 1975; Murray, 1967). As mentioned in Section 2.3.1, the operational basis for access to the phonological store is believed to differ for auditorily and visually presented memory items. Auditorily presented items are assumed to gain immediate access to the phonological store, while visually presented items have to be subvocalized first in order to be translated into an appropriate phonological code that enables access to the store.

The influence of long-term knowledge

One of the documentations of the beneficial effect of LTM on STM recall includes *the lexicality effect*, that is enhancement in serial recall accuracy for words as compared to nonwords (Hulme et al., 1991). Recall advantage explained in terms of LTM contribution is also found for high as compared to low frequency words (the word frequency effect, e.g. Gregg, Freedman & Smith, 1989), for lists presented in L1 as opposed to L2 (the language familiarity effect, e.g. Thorn et al., 2002), for nonwords containing highly frequent phoneme combinations (*the phonotactic frequency effect*) and for nonwords that are subjectively rated by listeners as resembling real words (*the wordlikeness effect*, e.g. Bailey & Hahn, 2001). Phonotactic frequency and wordlikeness ratings have been shown to be closely linked; nonwords that contain sequences with high-frequency phonological patterns tend to be rated as more wordlike

than those with low-frequency phonological patterns (Bailey & Hahn, 2001; Gathercole & Martin, 1996; Hay, Pierrehumbert & Beckman, 2003; Vitevitch et al., 1997).

The wordlikeness effect is referred to in the thesis on a number of occasions in relation to the characteristics of nonwords used in our own study and those used by other researchers. Wordlikeness can be defined as the degree to which a nonword resembles a real word. The degree of wordlikeness is normally obtained by asking native speakers of a given language to rate nonwords on a 5-point scale (“1” meaning “the nonword could not pass for a real word”). Nonwords of low wordlikeness are those that are of low relative similarity to existing words in a given language. Nonwords of high wordlikeness, on the other hand, highly resemble existing words. Obtaining wordlikeness ratings is part of the process of controlling the properties of nonwords. It has to be pointed out at this stage that it can be rather hard to *precisely* quantify the degree of wordlikeness of experimental nonwords (Smith, 2006: 585), as wordlikeness ratings are based on speakers’ *subjective* perceptions of sound sequence typicality. We acknowledge the fact that wordlikeness ratings are subjective in nature, but what speaks in their favour in our opinion is a strong statistical link with other more objective measures, such as e.g. phonotactic probabilities of nonwords (e.g. Gathercole & Martin, 1996; Vitevitch et al., 1997).

The wordlikeness effect (and other above-mentioned effects) can be explained in terms of the process of redintegration. The term *redintegration* (Schweickert, 1993) has been used to refer to a process whereby long-term memory representations influence short-term recall. Hulme and colleagues (1991) used the concept to explain why recall of words from short-term memory is better than recall of nonwords. What they propose is that in the case of word recall there are lexical representations available in LTM that could be used to repair partially decayed traces of words. They argue that in the case of nonwords no such representations are available, and as a result performance on nonword recall is lower than performance on word recall. The assumption that there are no lexical representations available for nonwords is incorporated in computational STM models for nonwords (e.g. Gupta & MacWhinney, 1997; Hartley & Houghton, 1996) presented later in this chapter (see Section 2.4.2). Gathercole et al. (1999a) extended the idea of redintegration to nonwords by illustrating that recall of nonwords of high wordlikeness is better than recall of nonwords of low

wordlikeness. Since recall of nonwords of high wordlikeness is affected by an individual's long-term knowledge of a language to a greater extent than recall of nonwords of low wordlikeness, recall of the latter type of nonwords (i.e. low-wordlike) is considered to be a purer representation of PSTM than recall of the former type of nonwords (Gathercole & Pickering, 1999).

2.3.3. Limitations of the Baddeley and Hitch Working Memory model – Nairne's critique

Although the concept of fractionated WM has been very prominent in publications in the last 30 years, within the WM tradition there is also much interest in a unitary view of memory and dynamic models of WM (see e.g. a volume edited by Osaka, Logie & D'Esposito, 2007).

A relatively recent critique of the concept of WM was undertaken in Nairne (2002). Nairne's criticisms are levelled at what he calls the 'standard model'. In the standard model short-term storage arises from activation, i.e. keeping information in an immediately accessible form. LTM is also activated and interacts with STM. Activation in STM is assumed to be fragile and can be easily lost via the operation of decay. Decay can be counteracted through the operation of rehearsal which refreshes activation. Baddeley's (2007: 37) comment on Nairne's conceptualisation of the 'standard model' is that, even though this resembles aspects of the Baddeley and Hitch WM model, it should not be treated as an adequate representation of the WM model and its phonological loop component.

In his critique Nairne raises a series of problems concerning the concept of subvocal rehearsal (an integral part of the WM model), to which Baddeley responds in his 2007 book (Chapter 3). According to Nairne (2002: 58), the account of rehearsal is undermined by the fact that, when rehearsal rates are controlled, individual differences in recall emerge (Schweickert et al., 1990). The reasoning behind this is that in the simplest form of the standard model most variability in immediate recall is attributable to rehearsal. However, as Baddeley points out (2007: 36), in the WM model both the capacity of the phonological store and rehearsal rates are expected to vary among individuals, resulting in individual differences in recall performance. Another related aspect that undermines the account of rehearsal, according to Nairne, is the finding that,

when rehearsal is controlled, different memory items lead to differences in recall performance (Hulme et al., 1991). For example, in Hulme et al. (1991) articulation rates for nonwords were slightly higher than for words, and yet words produced higher memory spans than nonwords (the lexicality effect). Baddeley (2007: 36) agrees that neither the lexicality effect nor the phonological similarity effect (i.e. better recall of phonologically dissimilar items as compared to recall of similar items) is likely to be dependent on the articulatory rehearsal process, and that a more complete model is needed, such as that provided by the phonological loop hypothesis. Nairne also points out that patients who have lost their ability to articulate speech may still show word length effects (i.e. higher recall accuracy for lists of short words as compared to that for lists of long words). Baddeley (2007: 36) explains this in terms of the fact that rehearsal involves the running of an inner speech programme and not necessarily overt articulation (Baddeley & Wilson, 1985). In dyspraxia, in which such inner speech programmes cannot be constructed, the word length effect is abolished (Waters & Caplan, 1996)¹⁰.

2.4. Mechanisms for short-term and long-term learning in computational STM models

The Baddeley and Hitch WM model posits a very close link between the phonological loop component (PSTM) and long-term learning of words (e.g. Baddeley, 2000). The WM model, however, does not directly address the actual mechanisms responsible for the interaction between the phonological loop and LTM and those responsible for the representation of order information in the phonological loop (Gupta & MacWhinney, 1997; Hitch, Flude & Burgess, 2009). As Burgess and Hitch argue (2005: 536), in order to get a better insight into the above-mentioned processes the Baddeley and Hitch phonological loop model needs to be translated into a quantitative computational STM model. A number of such computational STM models are in existence, including the Houghton (1990) model, the Burgess and Hitch (1992 & 1999) network model, the Hartley and Houghton (1996) linguistically constrained model, and the Gupta and MacWhinney (1997) model.

¹⁰ For a more detailed account of Nairne's critique please consult his 2002 paper.

2.4.1. Approaches to the storage and retrieval of serial order

Before we proceed to the presentation of the mechanisms for short-term and long-term learning in the above-mentioned computational STM models, it is essential that two general approaches to the issue of storage and retrieval of serial order in computational modelling of STM are outlined. These two approaches include chaining and contextual models.

Chaining provides one of the oldest accounts of serial order behaviour, and extends back to Ebbinghaus (1885). It assumes that items are stored in memory and revived via a chain of associations. A list of items is viewed as a sequence of associated pairs, in which each item can be used to cue the retrieval of a subsequent item (e.g. Lewandowsky & Murdock, 1989). To use the sequence *123456* as an example: within the chaining approach the list is seen as a sequence of five associated pairs, namely *1-2, 2-3, 3-4, 4-5* and *5-6*. Each item is believed to be associated with items presented in an adjacent position and, therefore, when *1* is cued it automatically calls up *2*, which in turn calls up *3*, and so forth.

The simplicity of this chaining mechanism is intuitively appealing, but problems arise when repeated items are employed within a list. If items are dependent on preceding items in the way that the chaining approach stipulates, then correct recall of a list such as *351954* should be a rather difficult task, since *5* serves as a cue for more than one item – *1* and *4* (Lashley, 1951). Another serious weakness of the chaining approach is related to the fact that it assumes that once an error has been made the chain is broken and the subsequently recalled sequence will, as a result, be erroneous. Research into recall patterns shows that one of the most common types of recall error is transposition, where two items are switched and the remaining part of the sequence remains unaffected (e.g. *25714* and *27514*) (Burgess & Hitch, 1992; Page & Norris, 1998).

An alternative approach to serial order behaviour is referred to as a *contextual approach*, and it assumes that items are retrieved from memory via some form of context that provides the cue. In the simplest form of a contextual model, the ‘slot model’, information storage and retrieval is likened to the mechanisms used in standard serial processing computers (Conrad, 1965). Information is believed to be stored in

STM in a limited number of storage locations and subsequently read off in the order of presentation. Other contextual models posit that serial order is determined with reference to a marker, usually the initial or the final item (Page & Norris, 1998; Shiffrin & Cook, 1978), or with reference to some form of timing mechanism (e.g. Houghton, 1990). The primacy model (Page & Norris, 1998), for example, posits that serial order is recalled via a series of associations with the first item in the list, the strength of the associations decreasing the further the item is situated from the marker. The Shiffrin and Cook (1978) model posits that the first and the last items in a sequence are markers and that storage takes place via associations between nodes that represent these two items.

In attempting to illustrate memory processes involved in the learning of phonological sequences, the mechanism referred to as competitive queuing also needs to be outlined. This mechanism has been taken as the basis of many computational STM models (e.g. Burgess & Hitch, 1992; Gupta & MacWhinney, 1997; Hartley & Houghton, 1996; Houghton, 1990; Page & Norris, 1998). The *competitive queuing* (CQ) mechanism was first proposed by Grossberg (1978 & 1987) and subsequently developed by Houghton (1990). In CQ, nodes that represent items from the sequence are activated one at a time (Layer 1). After the point of initial activation the strength of activation of each node decays. To account for the learning of sequences, the CQ mechanism posits higher-level nodes (Layer 2) that emit a context signal at each time step. A start node and an end node in Layer 2 are assumed to have weighted connections to nodes in Layer 1. At the beginning of the presentation of a sequence, a start node is activated in Layer 2, and its weights to Layer 1 are adjusted via Hebbian learning (Hebb, 1961) at each time step. At the end of the presentation, an end node is activated in Layer 2 and its weights are also adjusted. At the beginning of recall, the start node has a much higher activation level than the end node, and therefore it takes control over activation delivered to Layer 1 nodes. The closer to the end of recall, the more active the end node becomes relative to the level activation of the start node. When the level of activation of the end node is higher than that of the start node, the end node takes over the role of delivering input to the Layer 1 nodes.

A central assumption of the CQ model is the existence of the so-called *competitive filter*. The competitive filter is believed to be associated with Layer 1. It is

often described as a winner-take-all network, as it allows only one Layer 1 node to fire at each time step. A Layer 1 node with the highest activation level is selected first, and when it has fired it is immediately inhibited to allow for the selection of the node with the next-highest activation level, and so forth. This mechanism is consistent with the manner in which short sequences of actions queue for performance in the prefrontal cortex; cells that represent forthcoming actions are active in parallel, with the most active becoming the first-to-be-performed and the second most active becoming the second-to-be-performed, etc. (see Bullock, 2004, for more detail).

It should be noted that the CQ model, as described in Houghton (1990), deals with sequences of phonemes rather than sequences of words/nonwords. When a new word is presented, the relevant phonemes are activated in a layer of nodes representing phonemes (the phoneme layer), and the most recently presented phoneme is the most highly activated one. Nodes in the phoneme layer are in one-to-one correspondence with nodes in the competitive layer. The most active phoneme node in the phoneme layer communicates with its competitive filter node via a hard-wired excitatory connection and ensures that it wins the competition with other nodes, and consequently it becomes the only phoneme node to be allowed to produce an output at that time step.

2.4.2. Word learning mechanisms in selected computational STM models

Even though the CQ model, as formulated in Houghton (1990), addresses the processes involved in learning new vocabulary, it is only a partial account. The Houghton (1990) model was designed to deal with monosyllabic words, and it does not address the processes involved in serial recall of words or nonwords. Additionally, the model has no means of differentiating between novel and familiar phonological sequences (Gupta & MacWhinney, 1997). These limitations have stimulated the development of a number of computational STM models with a view to providing a more complete account of the interaction between STM and word learning.

The Burgess and Hitch (1999) model

The first of these is the model proposed by Burgess and Hitch in 1992 and further modified in 1999 (see also Burgess & Hitch, 2006 and Hitch et al., 2009). This

model was formulated to provide a computational framework for the articulatory loop component of the original version of the Baddeley and Hitch (1974) Working Memory model¹¹. Burgess and Hitch (1999) conceptualise the phonological loop as a neural network that consists of two main components, one being a phonological/lexical store for item information and the other being a context-timing signal that is responsible for encoding the serial order of items. The model includes input and output phonemes, the item layer and a context-timing signal. There are two pathways between input and output phonemes. One is a direct connection between the two types of phonemes, while the other is indirect, in that it involves an ordering mechanism and item nodes. The direct pathway is assumed to be a bi-directional connection, while the other pathway is perceived as a unidirectional connection that includes noise and decay with the passage of time. In Figure 2.3. below, which presents the relationship between elements of this model, boxes denote layers of nodes, dashed lines denote modifiable connections, and the solid line denotes a pre-wired connection.

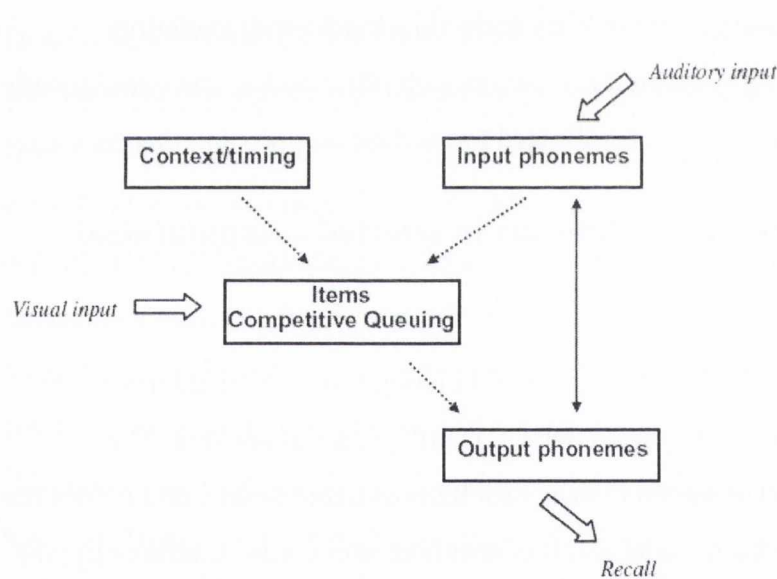


Figure 2.3. Outline of the Burgess and Hitch (1999) STM model

The Burgess and Hitch (1999) model extends the CQ framework formulated in Houghton (1990) to account for learning mechanisms in serial recall of words, and it addresses both serial order mechanisms in STM and the processes involved in long-term learning. In the model item nodes are linked to their corresponding input and

¹¹ The Burgess and Hitch (1992) model accounts for all the STM findings that the original 1974 WM model explained (e.g. the phonological similarity effect and the word length effect) and some additional STM findings that the original WM model failed to account for (e.g. transposition of adjacent items).

output speech representation and to a context-timing signal via connections that are capable of two kinds of modification, namely large short-term and smaller long-term modifications. During the presentation of an auditory item, its node is activated in the item layer, which triggers a learning process via strengthening connections between simultaneously active nodes in adjacent layers. This process of strengthening connections has two elements – one is large-amplitude and short-lived, while the other is small-amplitude and long-lived. Decay of the short-lived element is fast, which results in short-term forgetting. Decay of the long-lived element is slower, which allows long-term learning to accumulate when the item is repeated – provided that the repetitions are not spaced too far apart. During retrieval, item nodes are re-activated via their connections with the context-timing signal. Item nodes receive both phonemic and contextual information. The most active item node is then selected by the CQ mechanism and activates its output phoneme.

The way that rehearsal is represented in the Burgess and Hitch (1999) model has a direct correspondence to the way it is represented in the Working Memory model. Rehearsal is simulated in the Burgess and Hitch model as repeated retrievals, and it is used as a tool to refresh decaying short-term connections. During rehearsal input phonemes activate output phonemes. This corresponds to being able to hear one's inner voice while engaging in rehearsal. The Burgess and Hitch model also successfully simulates recall impairment resulting from articulatory suppression (i.e. when overt and covert recall is prevented by engaging in a concurrent verbal task). This is modelled as disrupting the creation of the input phoneme representation via a prior disruption of the output phoneme representation. During articulatory suppression, irrelevant output phonemes are activated, which leads to the activation of irrelevant input phonemes. As Gupta and MacWhinney (1997: 228) point out, the account of the rehearsal process is not entirely clear in the original version of the model. Rehearsal appears to be interpreted as making use of the chaining mechanism, and chaining models have difficulty accounting for some patterns of recall errors. Therefore, the subsequent 1999 version of the model abolishes the chaining assumptions.

Even though the revised version of the Burgess and Hitch model successfully deals with immediate memory for familiar phonological sequences (namely digits and words), it does not account for the error types observed in serial recall of lists of

unfamiliar phonological sequences (i.e. nonwords). If the phonological loop (PSTM) assists the storage and learning of novel phonological forms, as Baddeley and colleagues posit (see Sections 2.5 and 3.6), then the fact that the model does not distinguish between words and nonwords becomes a limitation. One of the attempts to remedy the problem is a model by Hartley and Houghton (1996) that proposes separate mechanisms for the recall of sequences at both sub-lexical and lexical levels.

The Hartley and Houghton (1996) model

The Hartley and Houghton (1996) model is referred to as a linguistically constrained STM model for nonwords, as it addresses the mechanisms involved in retention and immediate recall of nonwords within a framework of linguistic constraints. Word sequencing in this model involves a CQ mechanism similar to that of Burgess and Hitch, while nonword sequencing is posited to depend on a template or slot model that imposes constraints on the phonological structure in sequence production. Phonotactic constraints of the language, i.e. information about which phonemes can legally occupy a given position in a syllable in a particular language, are encoded by weights from the syllable template to the phoneme layer. These weights are permanent and are assumed to have been acquired as a result of previous experience (the solid line in Figure 2.4 below). The phonological form of syllables is encoded via two pathways, namely the content pathway and the structural pathway. The content pathway links the syllable to units that represent its constituent phonemes, while the structural pathway links nodes representing slots in the syllable temple to the nodes representing phonemes that can fill these slots. As the focus of the model is on immediate serial recall of nonwords, the binding of phonemes to syllable nodes is temporary (denoted as a dotted line in Figure 2.4).

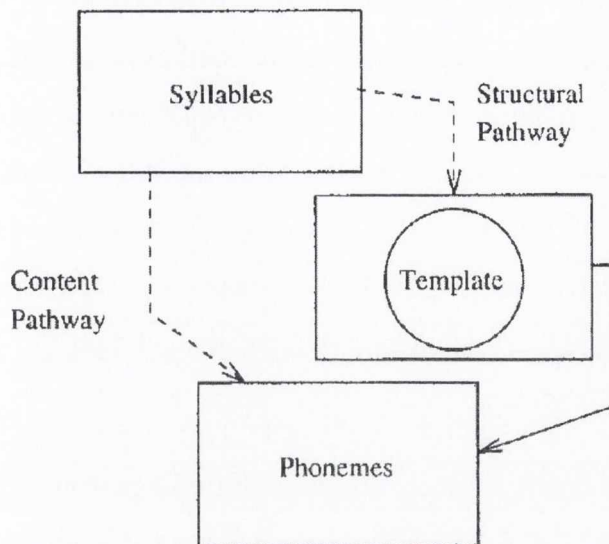


Figure 2.4. Outline of the Hartley and Houghton (1996) STM model

In the Hartley and Houghton (1996) model constraints on phonological structure are represented by two linguistic principles – the sonority rule and syllable structure. According to the former linguistic principle, in a well-formed syllable, sonority (i.e. the amount of energy in the speech signal) increases to a peak in the vowel and subsequently decreases (Selkirk, 1984). Each phoneme is assigned a sonority value, with obstruents and vowels having the lowest and highest sonority values respectively. According to the latter principle, syllables have an internal structure and are comprised of an onset (the initial consonant or consonant structure) and a rhyme (the vowel and the subsequent consonants) (Fudge, 1969). If a single node were to be used instead of the onset/rhyme division to represent a syllable, the sequences /mæt/ and /tæm/ would be represented identically in the syllable template. The onset/rhyme distinction, therefore, allows sequences that contain the same phonemes and occupy the same number of slots in the syllable template to be distinguished from each other.

During the presentation of a sequence of phonemes that represent a nonword, each phoneme that the model is presented with activates a node at the phoneme level. As mentioned earlier, the model establishes connections between the phoneme nodes and the syllable level via two pathways simultaneously – a direct (content) pathway to syllable units (encoding the phonetic content), and an indirect (structural) pathway via the syllable template nodes (encoding the syllabic structure). These two sequences of

input are bound by Hebbian adjustment of weights from the syllable node to a) the syllable template and b) the phoneme layer. During immediate serial recall of nonwords, the model recreates the serial pattern of activation of nodes in the network that occurred during the presentation stage, but it does so in reverse order (i.e. from the syllable layer to the phoneme layer directly or via the syllable template). As a result, a phoneme node receives input from both pathways simultaneously, and, when that happens, the item is then recalled. This is the process by which the system learns and recalls both the structure of a syllable and its constituent phonemes.

The Hartley & Houghton (1996) model accounts for research findings which illustrate that STM recall is influenced by linguistic mechanisms operating at a sub-syllabic level (see e.g. an earlier discussion of the influence of long-term knowledge on recall in Section 2.3.2). Since the model focuses mainly on explaining findings in relation to nonword recall, it does not distinguish between sequences that are nonsense syllables (e.g. *maf*) and sequences that are lexical items (e.g. *mat*). As a result, it does not account for the lexicality effect, i.e. increased memory span for words as compared to nonwords (Hulme et al., 1991). One way of addressing this issue would be to merge the Hartley & Houghton (1996) model for nonword recall with another STM model designed to account for word recall. The last model to be discussed in this section, the Gupta and MacWhinney (1997) model, does exactly that. It combines aspects of the Hartley & Houghton (1996) model with aspects of the Burgess and Hitch (1992) model.

The Gupta and MacWhinney (1997) model

The Gupta and MacWhinney (1997) model consists of the phoneme layer, the syllable template, the phonological chunk layer, the phonological store and the context maintenance queue (see Figure 2.5 below). The phoneme layer and the phonological chunk layer both have a CQ structure, i.e. they are composed of a set of nodes with an associated competitive filter. Nodes at the phonological chunk layer are a representation of the so-called chunks, namely either one-syllable word forms (e.g. *dog*) or multi-syllable word forms (e.g. *unwholesome*). The phonological store represents the element to which word forms can be temporally bound in sequential order, which is analogous to the item layer from the Burgess and Hitch (1992) model. The syllable template, placed between the phoneme layer and the phonological chunk

layer, imposes constraints on articulation that are analogous to the lexical constraints within the Hartley & Houghton (1996) model. The sequences of information that enter the phonological chunk layer are encoded both by the phonological store and the context maintenance (queue). Unlike encoding in the phonological store, which is automatic, context-maintenance encoding is a controlled, strategic process. The context-maintenance (queue) can be likened to the articulatory rehearsal process in the Baddeley and Hitch (1974) WM model (and in subsequent modified versions in Baddeley, 1986 & 2000).

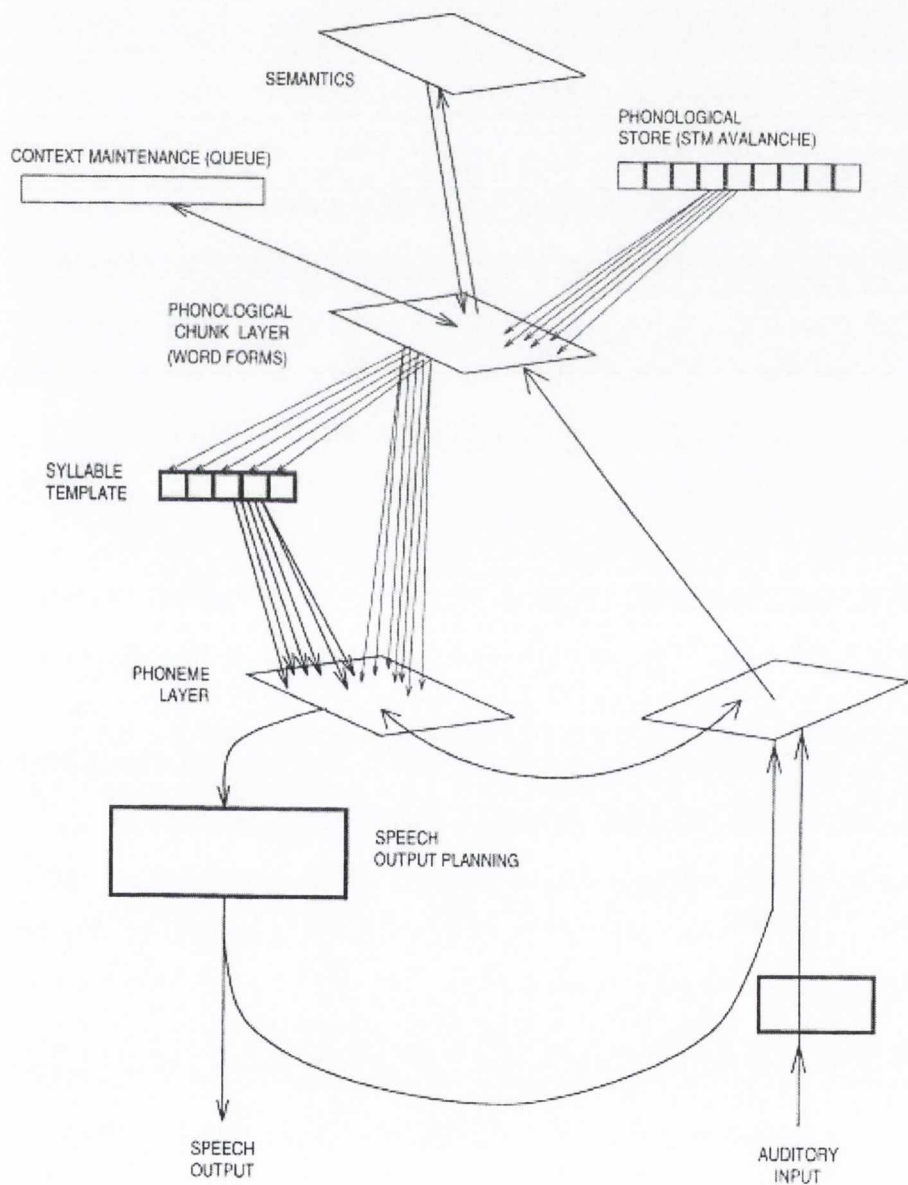


Figure 2.5. Gupta and MacWhinney (1997) STM model

Lexical and semantic information is represented as topologically organized in the Gupta and MacWhinney (1997) model. At the phonological chunk layer,

phonologically similar forms such as e.g. *cap* and *cat* are assumed to have neighbouring representations. In other words, they are organised according to a phonological similarity criterion. Phonologically similar items (e.g. *bat* and *rat*) contain similar information and, as a result, occupy slots that are closer together than dissimilar items (e.g. *bat* and *dog*). The phonological chunk layer nodes can receive activation from their associated semantic representations at the semantics layer, and can activate their associated semantics. The semantics layer is topologically organized in the same manner that the phonological chunk layer is, except that meaning is the criterion rather than phonological similarity.

In the Gupta and MacWhinney (1997) model, auditory information enters the system via two routes, one for familiar items (words) and the other for unfamiliar items (nonwords). When a familiar item is presented (e.g. *delighted*), its phonemes impinge on the phoneme layer in a sequential manner. Word recognition takes place and the phonological chunk node representing the relevant word is activated directly. When an unfamiliar item is presented (e.g. *rapuko*), the phonemes that make up the nonword are again activated one by one. Recognition processes operating on the nonword fail to activate any node at the phonological chunk layer, and a new chunk node representing the nonword is activated via the syllable template. The syllable template parses phonemes into syllables. The first cycle through the syllable template results in one syllable being processed. The phonological chunk layer has now weights attached to the phoneme layer representation of the syllable *ra*. The second cycle through the syllable template results in the phonological chunk node having weights attached to the phoneme layer representation of the two-syllable sequence *rapu*. The third cycle through the syllable template results in the phonological chunk node having weights attached to the three-syllable sequence *rapuko*. In this way a new phonological chunk representing a three-syllable nonword is activated.

Since the weights of this new phonological chunk are subject to decay, for the new word to be learnt permanently the effects of weight decay need to be reduced. During the process of cycling through the syllable template when a word/nonword is presented to the system, weights from the phonological chunk node to the constituent phonemes are automatically strengthened via Hebbian learning. When the presentation is over and the weights begin to decay, they can be strengthened via a re-presentation

of the nonword through rehearsal or another encounter (other encounters) with the word in the environment. This account of the role of rehearsal is consistent with the Baddeley and Hitch WM model. It could also be likened to the role that rehearsal is seen as playing in the Burgess and Hitch (1999) model when decay of a long-lived element is taking place.

During recall of a word or a nonword, the phonological store node is activated, which triggers the relevant phonological chunk node. The activation of the phonological chunk node sets up the activation gradient at the phoneme layer, which leads to sequential activation of the phonemes comprising the word/nonword. An output sequence of individual phonemes is subsequently produced via the operation of the competitive filter that is posited to be associated with the phoneme layer. It should be noted at this stage that in this model, the main difference between words and nonwords lies in the strength of their weights. Unlike in the case of words, when a nonword is first presented to the system no weights exist connecting a chunk node to the phoneme layer or to the semantic layer. These weights develop as part of the nonword (new word) learning process. Nonwords become words when their weights have become strengthened. In all other respects, words and nonwords are treated in the same manner. Both words and nonwords are represented at the same layer, namely the phonological chunk layer. This is in contrast with a suggestion made by Hartley and Houghton (1996) that words and nonwords may be represented at different layers.

In the Gupta and MacWhinney (1997) model, 'pure' serial recall ability that does not involve rehearsal is attributed to the phonological store, which corresponds to the Baddeley and Hitch (1974) WM model (and subsequent modified versions in Baddeley, 1986 & 2000). The existence of STM without rehearsal is documented in neuropsychological studies. For example, patient MK (described in Howard & Franklin, 1990), who had neurological impairment precluding rehearsal, was shown to have a certain (non-time-limited) auditory list memory capacity. The distinction between storage and rehearsal is also supported by neuropsychological evidence from patients that have lesions disrupting storage but not rehearsal (Vallar & Papagno, 2002). Additionally, neuroimaging research has located the phonological storage component in the temporo-parietal region in the left hemisphere. The rehearsal

component has been located more frontally in Broca's area (Jonides et al., 1998; Paulesu, Frith & Frackowiak, 1993).

For lists that exceed the capacity of the phonological store (usually more than five/six items, i.e. digits or one-syllable nonwords), normal adult subjects have to employ rehearsal strategies. As the list length increases, such strategic factors play an increasingly important role. When the list length exceeds adult subjects' phonological store capacity, such subjects are found to engage in cumulative rehearsal, a strategy that relies on subvocal rehearsal. Cumulative rehearsal involves attempting to create a chunk encompassing all the memory items and updating the sequence as each item is presented. Occasionally subjects resort to using other strategies, such as visualization (reported in the context of digit recall) or other mnemonic techniques (e.g. associating nonwords with similar-sounding words encoded in LTM). Mechanisms employed in cumulative rehearsal are incorporated in The Gupta and MacWhinney (1997) model.

The result of merging the Hartley and Houghton (1996) linguistically constrained model of STM for nonwords with the Burgess and Hitch (1992) STM model is the emergence of a new STM model that is capable of explaining the effects of LTM on STM performance and deals with both word and nonword recall. Without the incorporation of a mechanism that is capable of imposing a syllable structure and language-specific constraints on the order of phonemes, it would be rather difficult for a STM model such as that in e.g. Burgess and Hitch (1992 & 1999) to account for research findings indicating LTM influence on STM performance. By way of improvement, Nimmo (2004) suggests that the Gupta and MacWhinney (1997) model would benefit from incorporating the Hartley and Houghton (1996) representation of each syllable in terms of distinct onset-rhyme nodes. She justifies that viewpoint by arguing that without the distinct onset-rhyme representation the model would show large numbers of pre- and post-vocalic transpositions, which is inconsistent with previous research (Nimmo, 2004: 59).

The models described in this section constitute only a small fraction of all computational STM models attempting to account for recall performance. The reason why these particular models were selected is the fact that they address phonological sequence learning mechanisms within mathematically modelled versions of the phonological loop (PSTM). Another important reason for this selection is that the

relevant models can be presented in a way that is accessible to the reader who does not specialise in mathematics and computational modelling.

2.5. Brief overview of the functions of PSTM

This section provides a brief account of the functions ascribed to PSTM in respect of language learning within the WM framework. Please note that a detailed account can be found in Chapter 3, Section 3.6 (dealing with PSTM and the knowledge/learning of vocabulary) and Chapter 4, Section 4.6 (dealing with PSTM and the knowledge/learning of collocations).

The learning mechanisms posited in the computational STM models described in the preceding section provide an additional dimension for the subsequent overview of PSTM functions, giving more insight, for example, into why PSTM scores and L1/L2 vocabulary scores are linked in a number of correlational studies. As mentioned earlier, within the WM model no specific mechanism has been put forward to explain the processes involved in the link between the phonological loop (PSTM) and long-term learning of new words. Gupta and colleagues (Gupta, 2003; Gupta, 2005; Gupta & MacWhinney, 1997) suggest that both short-term and long-term learning of nonwords/new words involves representing sequences of phonemes within sequences of wordlike structures in memory. Individual differences in the ability to sequence phonemes and syllables could explain correlations obtained between nonword tasks and long-term learning of new lexical items (see Section 2.5.1 below). Since the present study also focuses on a possible link between individual differences in the ability to sequence wordlike units (elicited via either recall or recognition of sequences of nonwords) and long-term learning of combinations of lexical items (collocations), laboratory-based evidence regarding the learning of collocations is also presented below.

2.5.1. PSTM, L2 vocabulary and collocations, and the level of L2 proficiency

The relationship between PSTM and L1 vocabulary is particularly strong at lower levels of L1 proficiency. The diminishing role of PSTM in L1 vocabulary knowledge/learning as familiarity with L1 increases has been well documented

(Gathercole et al., 1992; Papagno & Vallar, 1995; Speciale et al., 2004: Experiment 1). This decrease of PSTM involvement with increased L1 proficiency level has led researchers to suggest that PSTM is responsible for long-term learning of *novel* word forms (Baddeley, Gathercole & Papagno, 1998; Gathercole et al., 1997). At the beginning stage of L1 learning, each word form that a child encounters is a completely *novel* word form. Gradually the child acquires more and more robust L1 phonological knowledge, on the basis of which increasingly robust phonological generalizations are made (Edwards, Beckman & Munson, 2004: 434). The more proficient the child is, the more robust his/her phonological generalizations become. At a higher level of L1 proficiency, L1 learners can access phonological representations of close neighbours of that new word. For example, while learning the adjective *unpleasant* they can access the adjective *pleasant* and the prefix *un* from LTM. This places a much lower burden on the limited-capacity phonological loop. This process is reflected in, for example, the Gupta and Tisdale (2009) model of STM, which simulates the interaction between PSTM, the existing lexicon and the learning of new word forms. The model is programmed in such a way as to enable learning phonological regularities from the input.

The picture presented in studies involving adult L2 learning is less conclusive than that emerging from developmental studies of children. The two existing classroom-based studies present mixed evidence. In the first longitudinal study by Speciale et al. (2004: Experiment 2) no significant link between PSTM and L2 vocabulary learning was detected in adult subjects at the onset of L2 learning. The second longitudinal study by O'Brien et al. (2006) suggests that a similar pattern to that observed in child L1 learning may be present in adult L2 learners in the classroom context, namely that the link between PSTM and vocabulary is stronger at lower levels of proficiency. In a group of adult L2 learners (the level of proficiency is not specified – possibly more advanced beginners or lower-intermediate students) the link between PSTM and productive use of vocabulary in spoken output was stronger in those subjects who were classified as lower ability students. Additionally, there are a number of laboratory-based studies involving adults which indicate that the link between PSTM and L2 vocabulary learning at the onset of L2 learning is very strong (Craik & Watkins, 1973; Gupta, 2003; Papagno & Vallar, 1992; Speciale et al., 2004, Experiment 1). In

order to contextualise the first of our research questions (Question 1¹²), a detailed overview of a number of studies dealing with the issue of PSTM and increasing vocabulary knowledge is presented in Section 3.6 in Chapter 3.

It would seem that a pattern of weakening of PSTM involvement should apply to adult L2 learning, as has been shown to happen in child L1 learning. After all if PSTM is a language learning device (Baddeley et al., 1998), its mechanisms of learning novel word forms should be applicable not only to the L1 context but also to the L2 context. It is possible, however, that the degree of PSTM involvement in adult L2 learning is affected by the fact that adults have another language system (their L1) at their disposal. Before they embark on the task of L2 learning, they already have extensive knowledge of L1 phonological regularities and a rich network of semantic information (see Chapter 3, Section 3.2.3 for a discussion of the bilingual mental lexicon). When there is a high degree of resemblance between L1 and L2 phonological regularities (e.g. Polish and Czech), adults may consult their L1 phonological representations while attempting to learn novel L2 words. In such a case, therefore, it would seem plausible to suppose that the impact of individual differences in PSTM capacity on vocabulary knowledge/learning could be weaker in that case than it is at initial stages of child L1 learning. When the degree of resemblance between L1 and L2 phonological regularities is lower (as is the case in Polish and English, i.e. the languages under scrutiny in the current study), L1 long-term knowledge may still be consulted, but the extent of such a consultation would be expected to be lower. The question, therefore, arises as to whether in the case of adult Polish learners of English the activation of L1 long-term knowledge is sufficient to compensate for individual differences in PSTM functioning, or whether the strength of the link between PSTM and vocabulary knowledge/learning is comparable to that detected in child L1 learning.

The Hartley and Houghton (1996) model and the Gupta and MacWhinney (1997) model (Section 2.4.2) do not address the issue of interaction between L1 and L2 long-term representations within their computationally modelled phonological loop. Both models include a syllable template that contains information about phonological constraints within one language (L1), and no suggestion is offered as to how another set

¹² **Question 1:** Does the relationship between PSTM and L2 vocabulary learning (in class and outside the classroom) decrease with an increased level of L2 proficiency in adults, as has been shown to happen in the case of child L1 learning?

of phonological regularities (L2) might interact with the modelled system. We believe that, in an exploration of the involvement of PSTM in adult L2 learning, L1 long-term vocabulary knowledge should also be included, and, therefore, this variable is part of our experimental design.

Apart from its role in long-term learning of novel word forms, PSTM may also play a role in the long-term learning of the order of novel word forms as well as the order of familiar word forms. The Gupta and MacWhinney (1997) model, for example, simulates the recall of both sequences of nonwords (novel word forms) and sequences of digits (familiar word forms). It seems plausible that that PSTM may be responsible for sequencing two-word collocations (such as those examined in this thesis) comprising two novel word forms (as in the Ellis & Sinclair study, 1996), two familiar word forms or a combination. The relationship between PSTM and the knowledge/learning of L2 collocations has not previously been explored in a classroom-based context. To the best of our knowledge, there is only one laboratory-based study that has addressed the issue of PSTM and L2 collocation learning in adults (Ellis & Sinclair, 1996). The study in question did not measure the subjects' variation in PSTM. Instead articulatory suppression was used in laboratory conditions to impair PSTM functioning. The articulatory suppression condition, which involves the use of a demanding secondary task that competes for the limited capacity of PSTM, was shown to produce a lower number of correctly remembered L2 phrases than other experimental conditions (i.e. the silent condition and the repetition condition).

Some indirect evidence suggesting that there may be a significant link between PSTM and L2 collocation learning in classroom contexts relates to work carried out in the area of L2 grammatical learning in adult learners. The O'Brien et al. (2006) classroom-based study showed that in their so-called low ability group Serial Nonword Recognition scores at Time 1 correlated significantly with free productive use of L2 vocabulary at Time 2 and L2 subordinate clause use at Time 2. In their high ability group, however, Serial Nonword Recognition correlated with L2 subordinate clause use at Time 2, but not with L2 vocabulary at Time 2. Accordingly, O'Brien et al. argue that at initial stages of language learning PSTM is mainly deployed for vocabulary learning, but that at later stages when vocabulary access is more automatic PSTM is redeployed for learning more complex structures (2006: 399).

The reason why both L2 vocabulary and L2 collocations are investigated in the course of the present study relates to the fact that we are interested in examining whether PSTM involvement in the learning of L2 collocations decreases at the same stage as PSTM involvement in the learning of L2 vocabulary (assuming of course that a decrease in PSTM involvement takes place in adult L2 learning as it does in child L1 learning). One possible explanation as to why there could be a difference in the degree of PSTM involvement in L2 vocabulary and L2 collocation knowledge/learning at a post-beginner level of L2 proficiency (B1 level in the current study) lies in the strength of links between syllables comprising a word as compared with the strength of links between words comprising a two-word collocation. In the Gupta and MacWhinney (1997) model, syllables are bound together into a word form via repeated cycles through the syllable template and are subsequently represented as a new phonological chunk layer node. Since the amount of exposure to a word is higher than the amount of exposure to a collocation in which that word features (see Chapter 3, Section 3.3 for a discussion of word frequency, and Chapter 4, Sections 4.3.2 and 4.5.2 for a discussion of collocation frequency), presumably it takes longer to strengthen links between words comprising a two-word collocation than it does to strengthen the links between syllables comprising a word. Additionally, when new phonological sequences (i.e. words) are learnt, generalised phonological knowledge is gradually acquired, which facilitates further vocabulary learning, and decreases the involvement of the phonological loop (as has been shown in child L1 learning). The order of syllables in a word is constrained by the information as to whether a given syllable is phonotactically legal in a given slot in a word. When it comes to learning two-word collocations, the way that words co-occur appears less predictable when compared to sequencing of syllables within a word. We speculate, therefore, that at the B1 level of L2 proficiency the involvement of PSTM in the learning of L2 collocations may be stronger than PSTM involvement in the learning of L2 vocabulary.

To contextualise another of our research questions (Question 4¹³) the issue of PSTM and the knowledge/learning of L2 collocations is discussed in more detail in Section 4.6 in Chapter 4.

¹³ **Question 4:** In the case of adults, is there a relationship between PSTM and the learning of L2 collocations (in class and outside the classroom) at the A2 and B1 levels of L2 proficiency? If so, does

2.5.2. The phonological store and the articulatory rehearsal component in L2 learning

Some researchers have voiced the need to compare recall-based tasks and recognition-based tasks (e.g. Baddeley et al., 2002) and their predictive power in L2 learning (Baddeley, 2007: 23; O'Brien et al., 2006: 395). Even though both Serial Nonword Recall and Serial Nonword Recognition test the retention of serial order information, speech-motor functions are involved only in the former case. Within the Working Memory model Serial Nonword Recognition is believed to mainly involve the phonological store (Baddeley, 2003a), whereas Serial Nonword Recall is assumed to involve both the store and the rehearsal component (Baddeley, 2003a; Bowey, 2001; Snowling, Chiat & Hulme, 1991).

Consequently, as indicated in Chapter 1, these two procedures may make different predictions in relation to L2 vocabulary and collocational learning (O'Brien et al., 2006: 395). In the case of young children the link between L1 vocabulary and a recognition-based nonword measure was shown to be stronger than that between L1 vocabulary and a recall-based PSTM measure (Gathercole et al., 1999b). The fact that both PSTM measures correlated with L1 vocabulary was interpreted as an indication that the link between PSTM and L1 vocabulary learning is related more to the phonological store component than the articulatory rehearsal component. However, an experimental study by Martin (2009) of adult artificial language learning, on the other hand, showed that in adult learners some aspects of L2 vocabulary learning and L2 word order were correlated with a recall-based measure but not with a recognition-based PSTM measure. This could indicate that in adults the involvement of the articulatory rehearsal component is stronger than it is in young children. This difference between young children and adults can be related to the finding that young children do not make use of rehearsal strategies in order to hold novel items in memory, while adults do (Gathercole & Hitch, 1993; Pickering, 2006: 245; Pressley & Van Meter, 1992).

the relationship between PSTM and L2 collocation learning decrease with an increased level of L2 proficiency?

Section 3.6 in Chapter 3 provides a more detailed background for two remaining research questions – Question 2¹⁴ and Question 5¹⁵.

2.5.3. The direction of causation in the link between PSTM and knowledge of L2 vocabulary and collocations

The issue of direction of causation the link between PSTM and vocabulary knowledge/learning has been a matter of intense debate in the last two decades. There are two alternative accounts, the PSTM account and the lexical account, that propose opposite directions of causality in the relationship between PSTM and vocabulary knowledge/learning. The PSTM account (Baddeley et al., 1998) posits that PSTM is a causal determinant of both Nonword Recall/Recognition ability and vocabulary knowledge/learning. Proponents of this view also acknowledge the impact of lexical/phonological knowledge on performance on nonword tasks (e.g. Baddeley, 2007; Gathercole, 1995; Gathercole et al., 1991). The opposing account, called the linguistic account (Snowling et al., 1991), proposes that it is lexical/phonological knowledge that causally determines performance on PSTM tasks. Both accounts acknowledge that the level of lexical/phonological knowledge may affect performance on nonword tasks. The major difference between the two accounts is that the PSTM account posits that PSTM plays an important *causal* role of in vocabulary knowledge and in the growth of vocabulary knowledge, while the linguistic account posits that the involvement of PSTM may be relatively *minor*.

The Gupta and Tisdale (2009) neural network model was developed to computationally addresses the issue of the interaction between PSTM and existing lexical knowledge in a longitudinal setup. The model offers an account of nonword repetition performance and novel word form learning (referred to by the authors as phonological vocabulary learning), as well as the interaction between the above variables. In short, the model consists of an input layer, and output layer, a hidden layer

¹⁴ **Question 2:** Which component of PSTM is more closely linked to L2 vocabulary learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?

¹⁵ **Question 5:** Which component of PSTM is more closely linked to L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?

and a context layer¹⁶. Each word form is presented to the model one syllable at a time, as it is computationally simpler to represent words as sequences of syllables than sequences of phonemes. After the word is presented, and when there is no longer any phonological information in the input layer, the model attempts to produce the word in the output layer. Over the course of many learning instances, the model learns new phonological sequences and also acquires generalised phonological knowledge.

Gupta and Tisdale (2009) carried a number of simulations within their neural network architecture in order to explore the opposition between the PSTM account and the linguistic account. Simulation 1 revealed that phonological vocabulary grows as a function of exposure in the simulated linguistic environment. Simulation 5 showed that variation in PSTM functionality affects vocabulary learning and that with no PSTM functionality no vocabulary growth takes place. Gupta and Tisdale interpret this as support for the PSTM account, which posits that PSTM is a critical and causal determinant of vocabulary learning. However, even when PSTM functionality is not explicitly manipulated, nonword repetition ability increases as a function of increasing vocabulary knowledge (Simulation 4). This could be interpreted as support for the PSTM account which acknowledges increased impact of lexical/phonological knowledge on performance on nonword tasks when vocabulary size increases. This could also *partially* be interpreted as support for the linguistic account which posits that lexical/phonological knowledge plays an important causal role in performance on nonword tasks. The model shows that this is indeed the case, but only at a higher level of ‘proficiency’, when vocabulary knowledge has expanded to a sufficient level. Notably, the Gupta and Tisdale simulations do not support the view that at the beginning of language learning the involvement of PSTM in vocabulary learning is relatively minor, as the linguistic account stipulates.

The discussion on the direction of causation is expanded in Chapter 3, Section 3.6 in order to contextualise two remaining questions in the current study – Question 3¹⁷ and Question 6¹⁸.

¹⁶ For a detailed description of the architecture of this model please consult Gupta and Tisdale (2009).

¹⁷ **Question 3:** In the case of adults (as opposed to children who were the focus of previous research), what is the causal direction of the relationship between PSTM and the learning of high-frequency L2 vocabulary (outside the classroom) at the A2 and B1 levels of L2 proficiency? Is it the L1-based nonword scores at Time 1 that contribute to L2 Vocabulary Levels Test scores at Time 2, or is it L2 Vocabulary Levels Test scores at Time 1 that contribute to the L1-based nonword scores at Time 2?

2.6. Selected PSTM measures: serial recall and recognition of nonwords

A number of procedures have been devised to measure the retention of a small amount of verbal material over a short period of time, and the most frequently used tasks include Nonword Repetition, Serial Nonword Recall, Serial Nonword Recognition, Serial Recall of Digits and Digit Span¹⁹. The PSTM tasks discussed in this section include those that are directly relevant to this thesis, namely Serial Nonword Recall and Serial Nonword Recognition²⁰. Serial Nonword Recall and Serial Nonword Recognition formats were selected for the present project, as one of our objectives was to compare results obtained via recall- and recognition-based tasks measuring STM for order information at a lexical level. Nonword Repetition was not included in the current study, as it measures memory for order information at a sub-lexical level. PSTM measures including digits and words were not included either, because they are known to be more susceptible to LTM influences (Hulme et al., 1991).

Serial Nonword Recognition differs from Serial Nonword Recall in that it measures individuals' memory for order information at a lexical level independent of their overt speech production ability. Serial Nonword Recall involves repeating back an auditorily presented sequence of nonwords in its original order (Archibald & Gathercole, 2007; Martin et al., 1999; Nimmo, 2004), while Serial Nonword Recognition requires indicating whether nonwords in two auditorily presented lists were presented in the same or different order (Gathercole et al., 1999a; Gathercole & Pickering, 1999; Gathercole et al., 1999b; Thorn et al., 2002). The performance on Serial Nonword Recall is typically scored using a correct-in-position criterion (i.e. scored as correct if each nonword is correctly repeated and if it is in the right position). This requires the participants to remember each item and the order in which they were presented. The performance on Serial Nonword Recognition is assessed on the basis of the number of lists correctly recognised as the same or different.

¹⁸ **Question 6:** What is the causal direction of the relationship between PSTM and L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency? Is it PSTM scores at Time 1 that contribute to Gitsaki's Collocation Test scores at Time 2, or is it Gitsaki's Collocation Test scores at Time 1 that contribute to the PSTM scores at Time 2?

¹⁹ Please note that for ease of reading the names of PSTM tasks are capitalised throughout.

²⁰ For a detailed historical review of other PSTM measures please consult e.g. Richardson (2007).

According to Baddeley (2003a), serial recall of phonological information relies on the phonological store and the articulatory rehearsal process, while serial recognition of phonological information relies on the phonological store only. The phonological store is responsible for creating new phonological traces, and the rehearsal process supports the trace creation. The phonological store is believed to be unaffected by LTM influences, whereas the articulatory rehearsal process is posited to be influenced by the familiarity of phonological sequences. This view is based on the pattern of results obtained in a study by Gathercole et al. (2001) showing that enhanced recall of words over nonwords is present in the recall procedure but not in the recognition procedure. Baddeley interprets this to mean that LTM influences are active for serial recall (relying on the store and rehearsal), but not for serial recognition (relying on the store only). Apart from being relatively impervious to LTM influences (Gathercole et al., 2001), Serial Nonword Recognition has also been shown to be impervious to the effect of subjects' phonotactic knowledge (Gathercole & Pickering, 1999) and to the language of testing (Thorn et al., 2002).

Additionally, it can be argued that from a methodological perspective there are merits to using two PSTM measures instead of relying on one measure exclusively (Gathercole & Pickering, 1999; Service, 2008, personal communication). One of the merits stems from the fact that, contrary to the earliest writings on the subject, PSTM tasks do not provide a pure measure of the construct. Apart from memory functions PSTM tasks are also known to tap other processes, such as e.g. speech-motor output processing (Gathercole, 2006a: 528-531). One of the implications is that, for example some individuals could obtain substantially lower Serial Nonword Recall scores than their 'true' PSTM score as a result of some minor articulation problems. To ensure that the PSTM capacity of subjects with some (even minimal) output problems is not underestimated by using a recall measure only, the use of Serial Nonword Recognition alongside Serial Nonword Recall has been strongly recommended (Gathercole et al., 1999b: 66). In subjects who do not have any output problems, the two measures should be highly correlated (provided that nonwords of low wordlikeness²¹ are used in Serial Nonword Recall, as described below).

²¹ The term *wordlikeness* is used to depict the extent to which experimental nonwords resemble real words (Section 2.3.2). High-wordlike (or high probability) nonwords are those that are thought to

2.7. PSTM scores and the language of testing

One final point that ought to be discussed in the present chapter is the impact of the language of testing on measuring variation in PSTM capacity.

In developmental studies of child L1 learning nonwords based on children's L1 have predominantly been used. When L1-based nonwords were employed to tap PSTM in children, PSTM scores increased alongside growing L1 lexical/phonological knowledge (Gathercole et al., 1992). When L2-based nonwords were employed in studies of child L2 learning, the same pattern could be observed, i.e. PSTM scores increased alongside expanding L2 lexical/phonological knowledge (French & O'Brien, 2008). However, when neither L1-based nor L2-based nonwords were used in studies of child L2 learning, children's PSTM scores remained stable over time, which has been interpreted as an indication that PSTM tasks based on such nonwords are relatively pure measures of PSTM capacity (ibid.).

With regard to adult L2 learning, both L1- and L2-based nonwords have been employed (e.g. O'Brien et al., 2006; Speciale et al., 2004). L1-based nonwords have been shown to produce stable results over a period of time when presented in the Serial Nonword Recognition format (O'Brien et al., 2006). The fact that Serial Nonword Recognition task is relatively impervious to the language of testing (Thorn et al., 2002), however, may have contributed to the stability of L1-based Serial Nonword Recognition scores over time in the O'Brien et al. study. The question about which types of nonwords reflect PSTM capacity most accurately in adults, therefore, requires further investigation.

In the current study we opted for nonwords constructed on the basis of the subjects' L1 (Polish). Our reason behind employing L1-based nonwords was to establish whether PSTM scores based on recall or recognition of such nonwords at the beginning and end of the study were stable. We worked on the assumption that our subjects' L1 phonological vocabulary knowledge would not change substantially over the research period. We expected that PSTM measures using L1-based nonwords might

resemble existing words in a given language to a high degree. Low-wordlike (or low probability) nonwords are the opposite.

produce similar results to those presented by French and O'Brien (2008) in relation to neither L1-based nor L2-based nonwords in child L2 learning.

Chapter 3: Vocabulary Breadth and Memory in Adult Second Language Learners

3.1. Introduction

The present chapter consists of five sections, four of which deal with various aspects of word knowledge and vocabulary breadth testing. The remaining focuses on one of the two central themes of this thesis, namely the link between Phonological Short-term Memory (PSTM or the phonological loop) and vocabulary learning in adult second language (L2) learners.

Section 3.2 outlines different approaches to defining the word, such as e.g. definition as a lemma and definition as a word family, and discusses a number of aspects of word knowledge, e.g. knowledge of form, meaning and word connotations. The notion of vocabulary breadth is outlined here within the wider framework presented by Nation (2001b) and that proposed by Daller et al (2007). Since adult bilinguals are the subjects in the present study, this chapter also discusses the organisation of the bilingual mental lexicon.

The concept of word frequency and its application in vocabulary breadth testing is introduced in Section 3.3. Lemmas and word families are referred to again in the context of the general corpora and corpus-based frequency lists that were a point of reference in the creation of vocabulary breadth tests used in our project. Milton's (2006c & 2009) notion of plateau phenomenon within high frequency bands (also Section 3.3.3) and estimations of vocabulary breadth at different proficiency levels within the Common European Framework of Reference (CEFR; Section 3.5) are outlined in this chapter as well, as they played a role in our decisions about the high frequency bands to analyse in studying vocabulary breadth gains at elementary (A2) and pre-intermediate (B1) levels of English proficiency. A critical review is also presented of a number of vocabulary breadth measures that tap receptive knowledge of vocabulary and productive knowledge – including the measures used in our own study (Section 3.4).

The final section (3.6) returns to the concept of PSTM introduced and discussed in detail in Chapter 2. Early studies of the link between PSTM capacity and vocabulary learning are presented first. A detailed discussion of the issue of the causality of the link between PSTM, L2 vocabulary learning and different levels of L2 proficiency follows. In this section we propose that – in order to get a better insight into the relationship between these two measures at different levels of proficiency – an analysis of the impact of PSTM on the rate of L2 vocabulary learning should be accompanied by a comparison of the strength of correlations between these two factors in two possible causal directions. This kind of statistical procedure (cross-lagged correlational analysis) has been used in relation to L1/L2 vocabulary learning in children (French, 2003; Gathercole et al., 1992) but not in relation to L2 vocabulary learning in adults. The reasons why the pattern of results might differ in these two populations are outlined.

3.2. Defining vocabulary and word knowledge

The word is quite a difficult concept to define in both theoretical terms and for applied purposes (Read, 2000: 17). A number of specialised definitions of what counts as a word exist, and yet the term *word* is quite frequently used in the literature without clarifying to which of the more specialised definitions it is intended to refer. For the sake of clarity, therefore, specialised definitions of word units such as lemmas, lexemes, word families, tokens and types are introduced in Section 3.2.1. The terms *lemma*, *lexeme* and *word family* are an essential part of our subsequent discussion of language corpora and frequency lists (Section 3.3.2), while the terms *token* and *type* re-surface in our presentation of measures of free productive ability (Section 3.4.3). Aspects of knowing a word, such as for example the breadth and depth distinction, are dealt with in Section 3.2.2.

3.2.1. Approaches to defining the word

The first of the specialised definitions of a word unit to be introduced here is the term *lemma*. A lemma comprises word forms that are limited to one part of speech, namely a headword and its most frequent inflections. The lemma of the verb “serve”, for example, includes its verb forms such as “serves”, “served” and “serving”, but not the noun form “service”. Again, the lemma of “easy” includes “easier” and “easiest”, as

they remain adjectives, but does not include the adverbial form “easily”. The form that is normally chosen to be the headword is the least marked one, e.g. the word “good” functions as a headword but not its inflected form “better”. When a lemma is used as a basic counting unit in word frequency lists, the lists are referred to as lemmatized word lists. Lemmatized word lists, such as e.g. British National Corpus lists, provide part-of-speech information but do not take account of semantic distinctions within what is considered to be a lemma. Semantic aspects are represented by another word unit called a lexeme discussed below.

A *lexeme* is also limited to one part of speech, but unlike a lemma it refers to a lemmatized form with a distinct meaning. The lemma “fair” (adj), for example, can be subdivided into a number of lexemes, such as “fair” = right and reasonable”, “fair” = treating everyone equally, “fair” = quite large, “fair” = light skinned, “fair” = not windy or rainy or “fair” = average (Longman Dictionary of Contemporary English Online²²). In dictionaries lemmas are used as headwords, while lexemes are listed as separate senses under a single entry. Counts of separate meanings of a word are used to create lexeme-based frequency lists, such as the General Service List (Section 3.3.2). Such lists are relatively rare, as the process of semantic tagging of a large general corpus remains something of a challenge (e.g. McEnery et al., 2006: 38).

The term *word family* reflects a broader approach to a word unit. A word family includes a group of words that share a common base to which different prefixes and suffixed are added. Bauer and Nation (1993: 253) define a word family as “a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately”. There are no hard and fast rules as to which types of words are included and which are left out, but English words are generally considered to be members of the same word family if they are inflected and derived items that contain affixes from the first six levels of the Bauer and Nation (1993) list. In their influential paper outlining a systematic approach to the classification of words into word families Bauer and Nation divide affixes into nine bands by frequency and productivity. At Level 1, for example, each running word is counted as a separate word. At Level 2 words with the same base and most common inflections are counted as separate units. This level broadly corresponds to the rules applied during the

²² The dictionary is available at <http://www.ldoceonline.com/>.

lemmatisation of the Brown Corpus (Section 3.3.2). Each subsequent level broadens the scope of affixes. To use the word “govern” as an illustration, forms that would be considered a lemma include “governs”, “governed” and “governing”, while forms that would be considered a word family additionally include e.g. “governable”, “government”, “ex-governor” and “governance”.

Having discussed such specialised definitions of a word as lemmas, lexemes and word families, we shall now move on to a distinction between tokens and types. The terms *token* and *type* are used to distinguish between the overall number of words produced and the number of different words produced. Calculating the number of tokens involves counting all running words in a written or spoken sample. Every individual word that occurs more than once is counted each time it occurs. Calculating types, on the other hand, involves totalling the number of different words used. Hence, a word is counted only once irrespective of repetition. To illustrate the difference between tokens and types, the sentence “The girl spoke to the teacher” contains six tokens but only five types, as the definite article “the” is repeated. While counting the number of tokens, reduced forms such as e.g. “aren’t” or “gonna” are separated into their constituent words. “Aren’t” is split into “are” and the reduced form of “not” (“n’t”), and “gonna” is split into “gon-” (a reduced form of “going”) and “-na” (a variant of “to”). In lemmatised lists these reduced forms are treated as variants of their full forms. The number of tokens and types is normally calculated to compare their relative proportions in native and non-native texts. The relative proportions of types and tokens will be dealt with in more detail in Section 3.4.3.

3.2.2. Breadth, depth and other aspects of word knowledge

One common convention is to divide word knowledge into breadth of word knowledge and depth of word knowledge, as suggested by Anderson and Freebody (1981). Vocabulary breadth (or size) is defined as the number of words for which the learners have at least some knowledge of their meaning. Vocabulary depth, on the other hand, focuses on the quality of knowledge of individual words, i.e. their collocational and grammatical patterns, associations, etc. The distinction between vocabulary breadth and depth is widely accepted (e.g. Goulden, Nation & Read, 1990; Lars, 2009; Shen, 2008; Wesche & Paribakht, 1996), but obviously word knowledge is too multidimensional a phenomenon to be fully tapped by this binary division. Researchers

accept that binary divisions like vocabulary breadth and depth are useful when focusing on specific aspects of lexical knowledge but many feel that they “do not really do justice to the complexity of word knowledge” (Milton, 2009: 14). To present a wider picture of vocabulary knowledge, we shall briefly outline two other approaches to knowing a word that incorporate the concepts of vocabulary breadth and depth but also include other dimensions. One of these approaches was developed by Nation (2001b) and the other by Daller et al. (2007).

In an attempt to provide a thorough summary of various types of word knowledge, Nation (2001b) addresses three broad areas, namely word form, word meaning and word use (see Table 3.1 below). Knowledge of word form involves being able to recognize and produce the written and the phonological form of a word. This type of knowledge also involves being able to recognize and use prefixes and suffixes that affect word meaning (see the previous section on lemmas and word families). The second type of word knowledge, i.e. word meaning, is quite often thought of as the ability to link the form to its meaning in particular contexts in which it has occurred. Apart from this most obvious aspect, knowledge of word meaning also includes being aware of associations carried by a given word. To use the adjective “fat” as an illustration, native speakers of English are aware that it carries rather negative connotations when used to describe a person. The same adjective, however, has had positive connotations in some parts of southern Nigeria, where women used to go to fattening houses to gain weight in order to look more attractive (Milton, 2009: 14). The final type of word knowledge, namely word use, concerns grammatical and collocational patterns in which the word occurs. As discussed in Chapter 3, some words commonly occur alongside some other words. For example, the adjective “high” collocates with the noun “mountain” or “cheekbones”, but not with the noun “flower”.

Table 3.1. Types of word knowledge according to Nation (2001b: 27)

Form	Spoken	Receptive Productive	What does the word sound like? How is the word pronounced?
	Written	Receptive Productive	What does the word look like? How is the word written or spelled?
	Word parts	Receptive Productive	What parts are recognisable in this word? What word parts are needed to express meaning?
Meaning	Form and meaning	Receptive Productive	What meaning does this word form signal? What word form can be used to express this meaning?
	Concepts and referents	Receptive Productive	What is included in the concept? What items can this concept refer to?
	Associations	Receptive Productive	What other words does this word make us think of? What other words could we use instead of this one?
Use	Grammatical functions	Receptive Productive	In what patterns does this word occur? In what patterns must we use this word?
	Collocations	Receptive Productive	What words or types of word occur with this one? What words or types of words must we use with this one?
	Constraints in use	Receptive Productive	Where, when and how often would we meet this word? Where, when and how often can we use this word?

As can be seen in Table 3.1 above, the framework contains references to the receptive and the productive aspects of word knowledge, i.e. being able to recognise a word versus being able to produce it (see Section 3.4.1 for a discussion of this distinction). As for the breadth and depth division, even though there are no direct references to it in the framework, we can assume that vocabulary breadth would include the Form category and possibly the Form and Meaning sub-section of the Meaning category, while vocabulary depth would cover all the remaining sections.

Another approach to aspects of word knowledge is the concept of three-dimensional lexical space (Daller et al., 2007), in which breadth and depth are contrasted with the level of fluency. Breadth is defined here as “the number of words a learner knows regardless of how well he or she knows them” (ibid.: 7), which maps the Form and the Form and Meaning sub-section of Table 3.1. Depth is defined as the degree of familiarity with the word’s associations, referents, grammatical functions, collocations and constraints on its use. Fluency refers to the degree of automaticity with which the learners are able to access and use the words they know. Breadth, depth and fluency are represented as three axes according to which the learner’s vocabulary knowledge can be located within three-dimensional space:

Some learners may have large vocabularies but are very limited in the speed and ease with which they can recall these words (...). These learners ought to be placed well along the breadth axis but less far along the fluency or depth axes. Other learners may (...) possess comparatively few vocabulary resources but considerable fluency in calling these to mind and using them in

communication. These learners would [be placed] less far along the breadth axis but further along the fluency axis.” (Daller et al., 2007: 9)

Daller et al. (ibid.) stress that the notion of lexical space is not meant to be treated as a fully developed theoretical framework, but rather as a metaphor attempting to depict the multifaceted nature of word knowledge.

3.2.3. The bilingual mental lexicon

Discussion of the nature of word knowledge reaches an even deeper level of complexity when the notion of bilingual mental lexicon is examined.

The word *lexicon* derives from a Greek word that means ‘dictionary’ and is used to refer to all the lexical elements present and operating in an individual’s mind. As indicated earlier, defining a word entails considering its numerous dimensions. Words comprising the mental lexicon can be operationalised in a number of ways, for example as lemmas, lexemes, tokens or types. They can also be conceived of in terms of different sets of characteristics, e.g. phonological, semantic or grammatical. Knowing a word involves knowing its phonological form, knowing its written form (in literate societies), and knowing other words it co-occurs with, as well as knowing about its grammatical profile. The mental lexicon is, therefore, conceived of as a more complex structure than that of an inventory of words and meanings. Its structure also reflects collocational profiles of word and their stylistic distribution.

The most widely-known model of a bilingual mental lexicon is that proposed by Weinreich (1953). According to this model, bilingualism can be classified into ‘subordinate’, ‘compound’ and ‘co-ordinate’ categories. In *subordinate bilingualism*, L2 words are seen as connected to the L1 mental lexicon in a parasitic fashion. The L2 system, on this view, strongly involves the L1 system, meaning being accessed and produced via the L1. In *compound bilingualism*, L1 word forms and L2 word form are seen as connected to one, fused conceptual system. A particular concept internalised by a compound bilingual, therefore, according to this account, has a single representation, but two word forms attached to it (L1 and L2). In *co-ordinate bilingualism*, two separate systems for of form-meaning links are posited to exist, so that an L1 word form and a corresponding L2 word for the same notion are viewed as having links to

separate semantic representations. Weinreich associates the above types of bilingualism with different types of learning experience, for example subordinate bilingualism is linked to learning the L2 in an L1-based context (e.g. a grammar-translation instructional environment). Weinreich acknowledges, however, that, within a single bilingual mind, different types of the relationship between L1 and L2 may co-exist. In more recent times the idea of the co-existence of a mixture of relationships has been explored further. For example, De Groot (1993, 1995) suggests that a bilingual's concrete words and perceived cognates may be stored across languages in a compound manner, while abstract words and non-cognates are more likely to be stored in a co-ordinate manner.

Lexical organisation in a bilingual mind, apart from being affected by the context of the learning experience, and by inherent properties of words, has also been argued to be affected by the level of L2 proficiency. De Groot and colleagues (De Groot, 1995; Kroll & De Groot, 1997; Kroll & Tokowicz, 2001), for example, refer to studies which indicate that subordinate structure in the bilingual mental lexicon is associated with lower levels of L2 proficiency, while the compound structure is associated with higher levels of L2 proficiency (see also e.g. Woutersen, 1997). Even though the division into subordinate, compound and co-ordinate bilingualism is intuitively appealing, it should be noted that such types of bilingualism are unlikely to be found in their 'pure' form (Cieřlicka, 2000). Cieřlicka posits that there are conceptual and formal-associative links between the L1 and L2 mental lexicons, and that nodes in the lexicons are connected by links that "vary in strength according to the type of a bilingual person's experience in his or her L2" (ibid.2000: 33).

With regard to the interaction between L1 and L2 phonological representations in the bilingual mental lexicon, it is posited that the acquisition of L2 speech sounds depends on the state of development of the L1 mental lexicon and on perceived cross-language phonetic distance (Flege & MacKay, 2004). At the beginning stages of L2 learning, L1 and L2 are believed to share a common phonological space. At that stage, it is claimed, the L1 acts as a template that influences the process of learning L2 phonological representations, and this is particularly applicable to adult L2 learners, since their phonetic categories are relatively well formed (McAllister, Flege & Piske, 2002). At the onset of L2 learning, according to this account, the L1-based template

containing information about L1 phonological constraints is consulted in the process of the creation of L2 phonological representations. As knowledge of L2 phonological representations becomes increasingly robust, it then becomes gradually more involved in facilitating the process of learning L2 phonological representations. These L1- and L2-based templates are reminiscent of a syllable template from the Gupta and MacWhinney (1997) model and the Hartley and Houghton (1996) model (described in Chapter 2, Section 2.4).

Considering the fact that L1 long-term knowledge may be linked to the process of the creation of L2 phonological representations in our adult subjects (at least at the A2 level), we decided to include a measure of L1 low-frequency vocabulary in our experimental design.

3.3. Frequency and vocabulary learning

Coming back to the notion of vocabulary breadth, it should be pointed out that the measurement of vocabulary breadth of L1 and L2 speakers quite often takes account of differences in word frequency. In the case of L1 speakers, such measurement involves tapping knowledge of highly frequent to highly infrequent words in order to obtain an estimate of the total number of words (or words families) that the native speakers know. With lower proficiency L2 learners, like the subjects in the current study, vocabulary breadth is normally defined in terms of their knowledge of relatively high-frequency words, such as those included in the General Service List or in the most frequent bands from the British National Corpus (BNC) list.

Measuring vocabulary breadth is of particular relevance, as the current study focuses on vocabulary breadth gains within specified frequency bands. The measurement of vocabulary breadth may involve a number of different aspects requiring slightly different operational definitions of the construct. It may, for example, involve probing the passive recognition of word forms with no attempt to elicit the word meaning (e.g. the X-Lex Test, Section 3.4). It may require linking a word form with its meaning (e.g. the receptive Vocabulary Levels Test, Section 3.4). It may also involve a productive translation, a picture-naming or a fill-in-the-blank requirement (e.g. the productive Vocabulary Levels Test, Section 3.4). As for the exact words to be tested, different types of vocabulary may be targeted. For example, the measurement of

vocabulary breadth may involve probing knowledge of words taught at a specific language course, or words expected to be known by L2 users at different proficiency levels or by L1 users at different ages.

3.3.1. Differences in word frequency

Some words tend to appear very frequently in spoken and/or written language, while others are relatively infrequent. There are words such as e.g. “the”, “of” or “and” that occur millions of times in large corpora (i.e. collections of texts), and there are other words like “interference” or “executive” that occur with significantly lower frequency. To distinguish between different levels of word frequency, a number of frequency bands have been defined. The first 1,000 frequency band (or level) contains words with frequency order numbers from one to 1,000. The second 1,000 frequency band (also referred to as the 2,000 frequency band) lists words with frequency order numbers from 1,001 to 2,000. The third 1,000 frequency band (or the 3,000 frequency band) is composed of words with frequency order numbers from 2,001 to 3,000, etc. Even though, from a conceptual point of view this division is rather straightforward, errors in the presentation of frequency bands occasionally take place, e.g. Milton (2009: 24) implies that the 5,000 word band (the fifth 1,000 word band) contains words with frequency order numbers of 5,001 and higher, while in fact the order numbers from the 5,000 frequency band fall between 4,001 and 5,000.

Differences in word frequency can be quite significant between and within frequency bands. Frequency data from the BNC lists compiled by Kilgarriff (2006) show that the words from the first 1,000 frequency band are on average 11 times more frequent than the words from the second 1,000 frequency band, and 40 times more frequent than the word from the fifth 1,000 band (Milton, 2009: 23). The within-band differences are most striking within the highest frequency bands. The first 1,000 frequency band ranges from 6,187,267 to 10,071 occurrences, the second 1,000 frequency band from 10,071 to 4,510, while the third 1,000 band from 4,508 to 2,516.

Apart from differences in word frequency, higher and lower frequency bands tend to contain different types of words. Most of the words from the first 1,000 frequency band tend to be function or structure words that carry relatively little weight of meaning. Less frequent words are mainly content or lexical words that “appear to

carry a greater burden of meaning in any sentence” (Milton, *ibid.*). Table 3.2 below gives a few examples of words from the first, second and fifth 1,000 frequency bands along with their frequencies of occurrence and their frequency order in the BNC.

Table 3.2. Examples of words from the first, second and fifth 1,000 frequency bands and their occurrences in the BNC (from Kilgarriff, 2006)

Frequency band	Order in the BNC	BNC occurrences	Word
First 1,000	1	6,187,267	the (determiner)
	2	4,239,632	be (verb)
	3	3,093,444	of (preposition)
	4	2,687,863	and (conjunction)
	5	2,186,369	a (determiner)
Second 1,000	1087	9227	cold (adjective)
	1089	9225	edge (noun)
	1090	9214	opposition (noun)
	1091	9213	opinion (noun)
	1092	9211	drug (noun)
Fifth 1,000	4525	1386	amend (verb)
	4526	1385	soap (noun)
	4527	1384	ruling (noun)
	4528	1381	interference (noun)
	4529	1380	executive (adjective)

Since in the literature vocabulary is often referred to as high- and low-frequency, a question arises as to where a borderline between frequent and infrequent words should be placed. Obviously there is no absolute cut-off point at which a word stops being frequent and becomes infrequent. High-frequency words are generally viewed as a very small but important group of words that “cover a very large proportion of the running words in spoken and written texts and occur in all kinds of uses of the language” (Nation, 2001b: 13). Usually the 2,000 word band is recommended as the most suitable limit for high-frequency words (Laufer & Nation, 1999: 36; Nation, 2001b: 14; Nation & Hwang, 1995). Nation (2001a) outlines five criteria for deciding the boundary between high- and low-frequency words, including the overlap between different frequency lists, and all of them appear to confirm that the 1,000 and 2,000 word bands ought to be considered as high-frequency words. Bearing in mind the fact that high-frequency words accumulate in the first, second and possibly third 1,000 frequency bands (Nation, 2004), the group of low-frequency words has a considerably higher number of members than the group of high-frequency words. The low-frequency group, apart from containing many words of very low frequency that are hardly ever used such as e.g. “gibbous” or “laths”, also contains some words of moderate frequency that did not make it to the high-frequency lists, as well as some technical vocabulary and proper names.

3.3.2. English language corpora and frequency lists

Measuring vocabulary gains in different frequency bands is closely linked to the notion of language corpora and corpus-based word frequency lists. This section homes in on those English language corpora and frequency lists were used as a point of reference for the vocabulary breadth tests and other procedures employed in the current study. It may be noted in passing that in our research project productive and receptive vocabulary gains from the first, second and third 1,000 frequency bands were measured by the receptive and productive Vocabulary Levels Test (Laufer & Nation, 1999; Nation, 2001b; Schmitt et al., 2001) which was created with reference to the General Service List (GSL; West, 1953), Thorndike and Lorge's list (1944), and the Brown Corpus (Kučera & Francis, 1967).

English language corpora

The term "corpus" was initially used to refer to a relatively small collection of naturally occurring samples of language from sources such as TV programmes or magazines, and could consist of "anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study" (Hunston, 2002: 2). Modern language corpora differ significantly from the original paper-based ones in terms of their size, structure and data storage. Modern corpora tend to comprise much larger numbers of texts that have been sampled in a systematic way to provide as representative a picture as possible of speakers' experience of a language in one specific context or in a number of contexts. The use of Internet-based (or PC installable) corpora instead of paper-based collections significantly increases the amount of information that can be accessed within a very short period of time and increases the number of users that can access a given collection at a given point in time. Software packages used to process electronic corpora broaden the scope of data that can be extracted and increase the accuracy of such an extraction.

Corpora can serve a number of different purposes, two of which are particularly relevant to our research project. First, the frequency data derived from the texts that electronic corpora include can be used to create word frequency lists. Corpus-based frequency lists, which rank words according to their number of occurrences, have become a point of reference in the process of creation of vocabulary size tests (Section

3.4) and some general proficiency tests (e.g. the cloze test procedure). Second, apart from their employment in vocabulary size testing, general corpus frequency data can be used as a baseline against which non-native spoken or written texts can be compared. L2 learners' lexical and collocational profiles can be calculated on the basis of such frequency data. Other uses of corpora in applied linguistics include for example comparing aspects of different varieties of a language or different languages (for more information see e.g. Hunston, 2002).

The corpora from which the frequency data for the current study were derived include the Brown Corpus²³ and the British National Corpus. The Brown Corpus, the first modern corpus of American English, was compiled at Brown University using approximately 500 chunks of 2,000 words of written text. All the texts were produced in 1961 and were sampled from 15 categories such as e.g. press reviews, general fiction or religion. The corpus is grammatically tagged, i.e. each word has a part-of-speech label allocated to it. Grammatical tagging allows the user to limit searches to a particular part of speech, and enables one to produce word frequency lists that take account of the grammatical category of a word as opposed to just its orthographic status (see Hunston, 2002, and Biber et al., 1998, for a description of automatic tagging and parsing).

The British National Corpus (BNC) is a 100 million word collection of present-day spoken and written English. Like the Brown Corpus, the BNC is grammatically tagged, which allows its users to make more specific queries. The BNC aims to represent modern English, and its contents are indeed more modern than those of the Brown Corpus. Written texts included in the BNC originate from dates no earlier than 1960, while spoken data come from no earlier than 1991. Over 93 per cent of the BNC texts date from the period 1985-1994. The written part (90 per cent) includes extracts from newspapers, journals, academic books, university essays, and other types of texts. The spoken part (10 per cent) involves different kinds of conversations ranging from informal chats between people of different ages and social backgrounds to more formal and structured discussions such as business and government meetings.

²³ The Brown Corpus, along with a number of frequency lists described later in this chapter, was employed in the creation of two research instruments selected for the current study, i.e. the Vocabulary Levels Test and the Lexical Frequency Profile method.

As Leech et al. (2001: 1) point out, the BNC has a clear advantage over other general corpora, such as e. the Bank of English, owing to the fact that its compilers have put a considerable amount of effort into ensuring that different kinds of written and spoken data were sampled broadly in accordance with how the data are represented in everyday life. From that perspective the BNC can be considered a quite balanced and broadly representative corpus. In terms of the written-spoken ratio, however, the corpus gives much more prominence to written data. In their attempt to justify the imbalance Leech et al. (2001) point to the difficulties involved in transcribing speech:

...Although spoken language, as the primary channel of communication, should by rights be given more prominence than this, in practice this has not been possible, since it is a skilled and very time-consuming task to transcribe speech into the computer-readable orthographic text that can be processed to extract linguistic information. In view of this problem, these proportions were chosen as realistic targets which, given the size of the BNC, are also sufficiently large to be broadly representative. (Leech et al., 2001: 1)

This brings us to the issue of corpus representativeness, a matter of some importance when corpora are used as the basis for tests. According to a stratified approach to corpus design (according to which the Brown Corpus and the BNC were developed), a corpus is broadly representative of the language as a whole when it consists of equal amounts of data from a number of contexts in which the language is used (Biber et al., 1998: 248; Hunston, 2002: 28). This approach requires cataloguing different categories of texts that exist in a given language and then obtaining samples from each of them. A well-designed general corpus has to represent a wide variety of registers within the language, and corpus designers go to great lengths to ensure that this criterion is met (see, for example, a description of the BNC design process in Crowdy, 1993). To add a word of caution, regardless of the selected sampling method there is no way of calculating the *exact* number of text types and their *exact* proportions that would ensure that a general corpus is fully representative of a given language as a whole:

...‘being representative’ inevitably involves knowing what the character of the ‘whole’ is. Where the proportions of that character are unknowable, attempts to be representative tend to rest on little more than guesswork. (Hunston, 2002: 28)

In interpreting data from a corpus, therefore, it is essential to bear in mind that the corpus can only reveal what was originally included in its contents, and that all generalizations made on the basis of frequency data drawn from that particular corpus are in fact extrapolations. In interpreting results of frequency-based vocabulary tests one should also remember that even very large samples of language, such as those contained in a general corpus, may not be representative of an individual student's exposure to the language, especially in relation to low frequency vocabulary (Nagy & Anderson, 1984: 321). Students may encounter some low frequency words more often than corpus-based counts would indicate, or they may not encounter them at all. This is less of a problem when it comes to high frequency vocabulary, but of course there may be some discrepancies between corpus-based counts and the counts derived from the learner's actual exposure to high frequency words. In spite of the above limitations of a corpus-based approach, a well-designed corpus can be treated as a useful tool for investigating various linguistic issues, including aspects of L2 vocabulary knowledge (Biber et al., 1998: 250), provided that some level of caution is exercised when drawing conclusions based on such frequency data.

Lemmas and word families as counting units

Before compiling a corpus-based frequency list in relation to which a vocabulary breadth or other language tests can be developed, a basic unit of measurement has to be first decided upon. Two main conventions for defining word count types are used in frequency lists, and they include lemma counts (e.g. BNC lemmatised frequency lists by Leech et al., 2001) and word family counts (e.g. BNC word family lists by Nation, 2006). As outlined earlier, a lemma comprises a headword and its inflected forms that belong to the same part of speech as the headword. Words from the same word family have a wider range of inflections and derivations than those included in a lemma, as they are not required to remain the same part of speech. The exact affixes that English words from the same word family can take are specified in the Bauer and Nation (1993) list of affixes (see Section 3.2.1).

In vocabulary breadth testing it is essential to be careful on account of what kind of measurement unit is adopted, as different word counts yield divergent vocabulary breadth figures. Early estimates of the vocabulary size of native speakers, for example, used a dictionary count (Seashore & Eckerson, 1940), according to which

words such as “see”, “sees” and “seeing” were treated as different words and counted separately. The Seashore and Eckerson approach yielded considerably higher figures of vocabulary size than contemporary calculations based on the constructs of a lemma and a word family as a counting unit. Early estimations of vocabulary size in adult native speakers of English ranged from 100,000 to 200,000 words (Seashore & Eckerson, 1940: 33), while modern calculations measuring lexical knowledge on the basis of word families indicate that well-educated native speakers of English have a vocabulary of around 17,000 base words²⁴ (Goulden, Nation & Read, 1990: 356).

The approach adopted by Seashore and Eckernson (1940) is no longer used in vocabulary breadth research; lemmas and word families are believed to provide better coverage in vocabulary testing and are also thought to be more practical. Instead of having to choose words from hundreds of thousands of dictionary entries, a smaller number of lemmas and an even smaller number of word families can be reviewed. The better coverage argument relates to the existence of regularities and rules according to which words are inflected and derived in a language. When new members from a known word family are learnt, knowledge related to making plurals, past forms of verbs, etc. is activated. As Bauer and Nation (1993: 253) point out, once the base word or a derived word is known, the recognition of other words from the same word family should require very little effort owing to the fact that the meaning of the base in various derived forms and its meaning when it stands alone are closely related. It ought to be noted, though, that the knowledge of one member of a given family does not automatically guarantee familiarity with another member of the same word family, especially at lower levels of L2 proficiency. This has practical implications for choosing frequency lists against which receptive and productive vocabulary size figures can be calculated for L2 learners at different levels of L2 proficiency.

Vocabulary size tests such as the Vocabulary Levels Test (Laufer & Nation, 1999; Nation, 2001b; Schmitt et al., 2001) and the X-Lex Test (Meara & Milton, 2003) adopt the lemma-based definition of a word. Lemmas are generally recommended as a more valid unit of measurement with regard to L2 learners’ productive knowledge regardless of L2 proficiency (see e.g. Vermeer, 2004). If word families were to be the

²⁴ Despite the expectation that native speakers should have very large vocabularies, there appears to be fairly systematic variation around Goulden et al.’s figure. As Milton (2009: 246) reports, undergraduates entering university score around 9,000 on the Goulden et al. test, post-graduates score between 13,000 and 14,000, and faculty members over 20,000.

measuring unit, as they are in the measurement of vocabulary breadth in native speakers, L2 learners could get credit for some words from a given word family that they cannot produce, and they could also lose credit for not knowing the entire word family when they can in fact produce some words from the relevant family. In relation to L2 learners' receptive knowledge, lemmas are considered a valid type of a word count when dealing with lower proficiency learners:

... the learners at this level [i.e. elementary or intermediate] are likely to have mastered only the most frequent inflections and derivations, but will not know the more infrequent and irregular ways in which words can change. By using lemmatised wordlists as the basis for tests at this level, we get believable and stable results. (Milton, 2009: 11)

For more advanced learners word families are considered to be a more valid measurement unit than lemmas (Bertram, Baayen & Schreuder, 2000; Bertram, Laine & Virkkala, 2000; William Nagy et al., 1989). If lemmas were used as the basic unit of measurement, there would be a danger of learners getting higher scores at lower frequency levels by demonstrating the passive knowledge of lower-frequency lemmas from a higher-frequency word family (Nation, 2007: 39).

English language frequency lists

A word frequency list is a sorted list of words or word types such as lemmas or word families. Frequency lists normally provide the number of occurrences for lemmas and their members in the corpus from which the lists are drawn. A small number of lists also provide semantic information that allows the user to distinguish between homographs, i.e. words with distinct multiple meanings (the General Service List). Apart from providing frequency data for each lemma, frequency lists such as the lemmatised BNC lists (Leech et al., 2001) also provide range data (the number of subdivisions within the corpus in which the lemma occurred), dispersion data (similarity between the frequencies across the subdivisions) and part-of-speech information.

Before we present relevant frequency lists, it should be pointed out that there is no single frequency list for general English or particular uses of English that could be called a definitive word frequency list (Read, 2007: 109). Different lists have different

strengths and weaknesses. Some lists, for example, have the advantage of taking into account word meanings (e.g. the GSL), while others have the advantage of keeping up with language changes and incorporating newly created words (e.g. the BNC or the Bank of English). Therefore, “in drawing a sample of words for a vocabulary size test it is very much a case of using the best available (or perhaps, least unsatisfactory) list” (Read, 2007: *ibid.*).

The most influential of the early frequency lists is the General Service List (West, 1953). Despite the fact that it was created over half a century ago the GSL still provides the foundation for work in the field of L2 vocabulary knowledge (Laufer, 2005; Meara, 2005; Morris & Cobb, 2004; Muncie, 2002). The GSL is a grammatically and semantically tagged list composed of over 2,000 words²⁵. When re-counted as word families, the GSL includes 1,986 headwords (Nation, 2004: 7). The words selected for inclusion in the GSL were believed to be “most suitable to set as a first objective for foreign learners” (West, 1953: iv) and were mainly, but not completely, chosen on the basis of frequency. Other selection criteria included range, familiarity and pedagogical value. Since computer counts were not available at the time, West used hand-counted frequencies from Thorndike and Lorge (1944). Even though the GSL has been criticised as outdated (Coxhead, 2000; Engels, 1968), it accounts for as high a percentage of “the running words in any written or spoken English text, now as [it] did fifty years ago” (Read, 2007: 108). Additionally, it possesses a certain quality that many frequency lists based on modern large corpora such as the BNC do not possess, namely information about the frequency of the different meanings of words. This kind of information is believed to be essential in the process of creating high frequency word lists that are valid (Engels, 1968: 214; Graham, 2008). In recent years the GSL was adopted as a point of reference in the making of the Vocabulary Levels Test (Section 3.4).

There are also a number of lemmatised BNC lists. The first set of lemmatised BNC lists was developed by Leech et al. (2001). To create the lists Leech and his colleagues rearranged the BNC into 100 subdivisions, each consisting of one million running words from similar types of texts. Then they selected lemmas occurring in the corpus more than 1000 times, and provided frequency data for each lemma (number of

²⁵ The list is available at <http://jbauman.com/gsl.html>.

occurrences in the corpus), range data (the number of subdivisions in which the lemma occurred), dispersion data (the level of similarity between the frequencies across the subdivisions) and part-of-speech information. Other lemmatised lists compiled on the basis of the BNC include the work of Kilgariff (2006) and Chujo (2004). Kilgariff, a leading corpus lexicographer, created two versions of the BNC frequency lists, one lemmatised and the other one raw (not lemmatised). The lemmatised list contains 6,318 lemmas, while the raw list includes 208,657 types that occur over five times in the entire BNC corpus. The lists provide basic information such as frequency and part of speech, but unlike the lists by Leech et al. no dispersion or range data are given. The Kilgariff raw list was used by Chujo (2004) to compile yet another lemmatised BNC word list, longer than that by Kilgariff. The Chujo list contains over 14,000 lemmas representing 86,123,934 tokens in the BNC. The most recent set of the BNC lists was developed by Nation (2004 & 2006) on the basis of the BNC counts and the lemmatised BNC lists created by Leech et al. (2001). A word family, not a lemma, was selected to represent the basic unit of measurement.

A comparison of the high-frequency BNC lists (the first, second and third 1,000 word-family lists) with the GLS plus the AWL (Academic Word List; Coxhead, 2000 & 2002) reveals that, despite the fact that the two sets of lists were created at different times and on the basis of different corpora, overall they contain quite similar vocabulary (Nation, 2004). Even though there are some differences in terms of the distribution of the vocabulary within each frequency band of the two sets, the general tendency for high-frequency vocabulary from different general frequency lists is to remain largely the same (Nation & Hwang, 1995). As Nation (2001b) puts it, general frequency lists “may disagree with each other about the frequency rank order of particular words but if the research is based on well-designed corpus there is generally about 80% of agreement about what particular words should be included” (Nation, 2001b: 15-16).

3.3.3. The frequency model and vocabulary learning

The underlying principle behind the most commonly used vocabulary breadth tests is that learners, as a group, are sensitive to the frequency of words they encounter. The idea that word frequency affects the level of word difficulty is an old one and dates back to the early 20th century. As early as 1917 Palmer expressed the view that “the

more frequently used words will be the more easily learnt” (1917: 123), and by extension that low-frequency items will by large require more learning effort. This concept has since been picked up by a number of academics (e.g. Mackey, 1976; McCarthy, 1990; Nation, 1990), but until very recently there has not been much in the way of empirical evidence to support the frequency hypothesis (see e.g. Wesche & Paribakht, 1996).

Meara’s (1992) model of vocabulary learning has become a point of reference for many subsequent studies exploring the relationship between the frequency factor and vocabulary breadth (Milton, 2006a, 2006b & 2006c; Milton, 2009: 34; Richards & Malvern, 2007). As illustrated in Figure 3.1 below, Meara’s frequency profile of a typical L2 learner graphically represents the level of vocabulary knowledge from the first, second, third, fourth and fifth 1,000 frequency bands. According to the model a typical learner’s vocabulary knowledge would be expected to be higher in the more frequent columns and lower in the less frequent columns. As the learner’s proficiency increases, the knowledge of each band should ideally move upwards till it hits a ceiling of 100 per cent. Higher frequency bands are believed to reach this ceiling before lower bands begin to approach it.

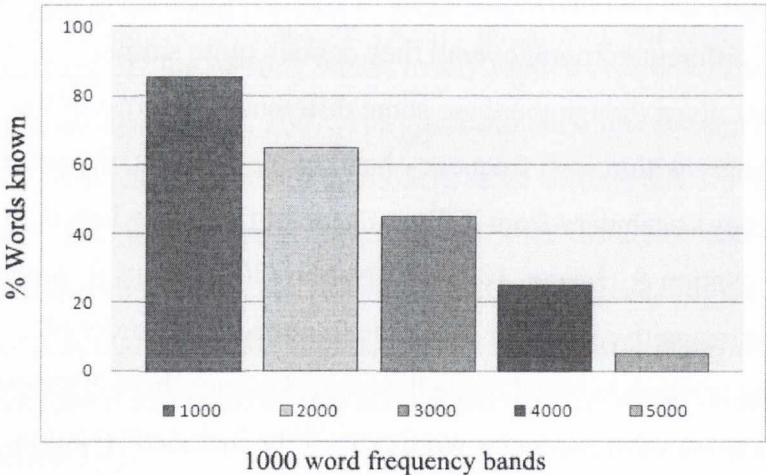


Figure 3.1. Frequency profile of a typical learner (Meara, 1992: 4)

In a recent study by Milton (2006c) the obtained group frequency profile resembles the model put forward by Meara in that the score in the first band is higher than that in the second band, which in turn is higher than that in the third band, etc. (see Figure 3.2 below). There is a clear downwards slope from left to right, but the slope is noticeably less steep than that presented in Meara’s (1992) graph of a typical frequency

profile. The learner's knowledge of vocabulary from the first 1,000 band exceeds their knowledge of words from the fifth 1,000 band by over 50 per cent.

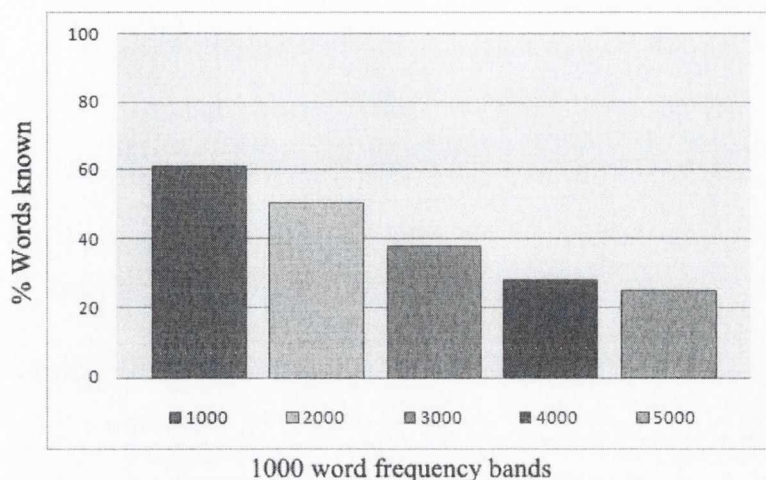


Figure 3.2. Frequency profile of Greek learners of EFL (Milton, 2006c)

Our decision about which frequency bands to examine in the course of the current study (A2 and B1 levels of proficiency) is partly based on a study by Milton (2009: 32) described below. The frequency model by Meara (1992) assumes that vocabulary knowledge of separate frequency bands eventually reaches something in the region of 100 per cent. The model implies a steady rise in vocabulary knowledge in each frequency band, with high-frequency vocabulary knowledge reaching the maximum first. In reality, however, the learning process is far less straightforward in that some frequency bands develop faster than others (Milton, 2006b: 199) and that the maximum is hardly ever reached (Milton, 2006c). To examine whether learners achieve 100 per cent in most frequent bands, Milton (2009: 32) drew up individual profiles of the 10 highest scoring Greek learners of English from his earlier study (Milton, 2006c). When the profiles are looked at, a rather high level is achieved in most of the frequency bands under scrutiny but hardly ever 100 per cent. As can be seen in Figure 3.3 below, only one student reaches 100 per cent in the 1,000 frequency band and two other students reach nearly 100 per cent in the 2,000 and 3,000 frequency bands. Milton's data suggest that the highest scoring students' frequency profile in the most frequent bands rises to up to approximately 85-90 per cent before it begins to plateau.

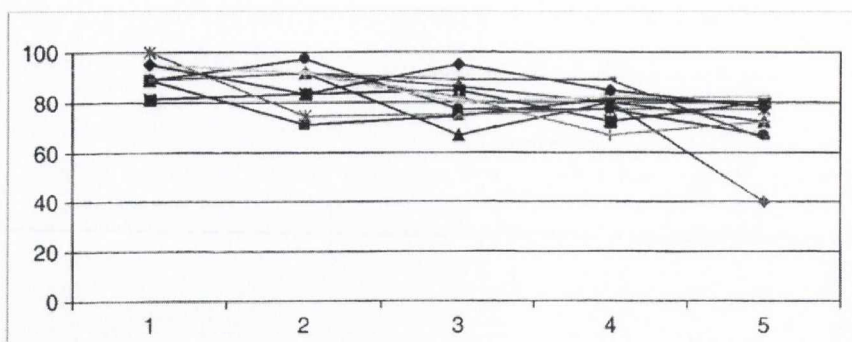


Figure 3.3. Individual frequency profiles of 10 highest scoring learners of EFL in Milton (2006c)

It is not quite clear yet why the profile plateaus at that point, but the plateau phenomenon should be taken into account in studies like ours, where for practical reasons only a small number of frequency bands can be examined over time and where the focus is on vocabulary gains rather than on proving or disproving Meara's frequency profile. In the frequency bands in which the students' knowledge on average is rather high, substantial gains are quite unlikely to be observed over time. For that reason, after an extensive pilot study the first 1,000 frequency band was excluded in examining vocabulary gains of our pre-intermediate (B1) students. Our decision about the frequency bands to be examined in the elementary (A2) and pre-intermediate (B1) groups was also partially motivated by research on vocabulary breadth and different proficiency levels within the Common European Framework (Section 3.5) and by our own pilot study (Chapter 5, Section 5.3.3).

3.4. Measuring vocabulary breadth

Having discussed the notion of frequency counts in L2 vocabulary learning, we shall now move on to various test formats that probe the knowledge of vocabulary from different frequency bands. This section presents some of the most widely used test formats for measuring receptive and productive vocabulary breadth, including the test employed in the current study (i.e. the receptive and productive Vocabulary Levels Test). Strengths and weaknesses of a number of vocabulary breadth procedures are outlined, to indicate the kind of factors taken into account in deciding which test format to employ in the present study.

3.4.1. The receptive versus productive distinction in vocabulary breadth

A number of different test formats have been used by researchers to determine the size of non-native speakers' receptive and productive vocabulary. In relation to the size of L2 receptive vocabulary, the test types that have been most widely used include a definition-matching format, such as the Vocabulary Levels Test (Nation, 1983), and a checklist format, such as the Eurocentres Vocabulary Size Test (Meara & Buxton, 1987; Meara & Jones, 1987). The size of productive vocabulary has been frequently measured by means of a fill-in-the-blank format (e.g. Laufer & Nation, 1999) and a translation format (e.g. Kudo, 2001). Some test formats, such as translation, have been used for assessing both receptive and productive vocabulary size (e.g. Laufer & Goldstein, 2004; Takala, 1984; Webb, 2008)²⁶.

Despite the fact that separate formats for receptive and productive vocabulary exist, the distinction between receptive and productive knowledge is not a straightforward one. Even though it is generally assumed that newly learnt words are initially available in their receptive dimension and that they become available for productive use at a later stage, the exact threshold at which a passively known word obtains a productive status would of course be rather difficult to locate (Read, 2000: 154). With regard to assessing production and reception, two general approaches can be identified – one focusing on recognition and recall, and the other measuring comprehension and use (Read, 2000: 154-155). Recognition is often depicted as the ability to recognize a word in its spoken or written form and to recall its meaning when presented with the word, while recall is defined as being able to provide the target word from memory as a response to some sort of stimulus, such as an illustration or a fill-in-the-blank prompt. *Comprehension* means understanding a word encountered in context while listening or reading, whereas *use* means that the word occurs in the learner's own speech or writing. The *recognition vs. recall* approach makes use of relatively context-independent test items, while the comprehension vs. use approach involves testing words via context-dependent tasks. Measures of vocabulary breadth discussed below belong to the former tradition.

²⁶ It would be interesting to look into how the vocabulary tests presented in this section relate to models of lexical access and retrieval. However, this is beyond the scope of the thesis.

3.4.2. Receptive measures of vocabulary breadth

In order to test receptive vocabulary breadth in as reliable a manner as possible, knowledge of a relatively high number of items has to be probed. A larger item pool is more likely to produce a accurate picture of overall vocabulary size than a smaller item pool (Milton, 2009: 72). For that reason test formats that contain quick-to-complete discrete and context-independent items are favoured in vocabulary breadth research over other more-time consuming formats.

One of the best-researched context-independent formats involves matching explanations of foreign words with the target words. This format has been adopted by Nation (1983; 1990; 2001b) and colleagues (Laufer & Nation, 1999; Schmitt, 2000; Schmitt et al., 2001) in their test of vocabulary breadth called the Vocabulary Levels Test or the Levels Test. The test format in question looks as follows:

- | | |
|--------------|-----------------------------------|
| A. bake | 1. join together (B) |
| B. connect | 2. walk without purpose (F) |
| C. inquire | 3. keep within a certain size (D) |
| D. limit | |
| E. recognize | |
| F. wander | |

The Vocabulary Levels Test has been used by many researchers to explore second language learners' receptive vocabulary size within specified word frequency limits (e.g. Beglar & Hunt, 1999; Cameron, 2002; Fan, 2000; Laufer & Paribakht, 1998; Schmitt et al., 2001). Apart from its widespread application in cross-sectional studies it has also been shown to be useful in assessing gains in vocabulary size over time (e.g. Tschirner, 2004; Xing, 2007). There are a number of parallel versions of the Levels Test (see a detailed description in Chapter 5, Section 5.3.3 and Chapter 6, Section 6.3.5), and the version employed in the current study is that developed by Nation (2001b) and modified by Schmitt et al. (2001).

The receptive Levels Test is based on the notion of frequency levels drawn from corpus-based frequency lists (Section 3.3). The Levels Test consists of five sections, each testing knowledge of 30 words. The word frequency levels measured via the definition-matching format include the following: 2,000, 3,000, 5,000 and 10,000 and the University word level (Xue & Nation, 1984). There is also a separate receptive

component for the 1,000-word frequency level (Nation, 2001)²⁷. This component differs from all the other components in that it has a True/False format (see Chapters 5 and 6). The most recent version of the Levels Test, the so-called Vocabulary Size Test (Nation & Beglar, 2007), samples knowledge of vocabulary from as many as 14 frequency bands, and consists of 10 multiple-choice items at each frequency level sampled from word-family BNC lists. The following test items come from the 10,000-word frequency level of the Vocabulary Size Test:

1. awe: They looked at the mountain with **awe**.
 - a) worry
 - b) interest
 - c) wonder
 - d) respect
2. peasantry: He did a lot for the **peasantry**.
 - a) local people
 - b) place of worship
 - c) businessmen's club
 - d) poor farmers

By totalling the scores from separate frequency sections of the Vocabulary Levels Test an estimate of overall vocabulary size can be arrived at (see different calculations in Gyllstad, 2005, Laufer, 1998 and Schmitt & Meara, 1997; 1997). It is not uncommon, however, for the results from different frequency sections to be treated as a profile of vocabulary size and not to be totalled into one overall figure (e.g. Cameron, 2002; Schmitt et al., 2004a). Whether or not the scores from separate frequency sections are totalled largely depends on the purpose of a given study.

Another very popular format in vocabulary size research is a checklist, which requires that the test-takers indicate the words they recognize. The test-takers are presented with a series of words and asked to tick the ones they think they know and can use. This format is often combined with a system that allows some kind of compensation for guesswork or overestimation of one's knowledge. Made-up words, i.e. words that do not really exist but read and sound like real words, are often included in such tests. While calculating the final score, the scores obtained for Yes responses to the made-up words are deducted from the raw score. The following extract is taken from the X-Lex Test (Meara & Milton, 2003), a test of receptive vocabulary size

²⁷ The 1,000-word frequency level means the first 1,000 words most frequently occurring in the English language. The 2,000-word frequency level, by analogy, means the second 1,000 words, etc.

(Figure 3.4). The first column on the right contains made-up words, while the other columns contain real English words from 1,000 to 5,000-word frequency levels. Words from the first, second, third, fourth and fifth 1,000 word bands are presented in columns 1, 2, 3, 4 and 5 respectively. The authors advise that during the administration of the X-Lex Vocabulary Test the words should be presented in random order.

X-Lex Vocabulary Test					
Please look at these words. Some of these words are real English words and some are invented but are made to look like real words. Please tick the words that you know or can use.					
dog ✓					
that	both	cliff	sandy	lessen	darrock
with	century	stream	military	oak	waygood
before	cup	normal	impress	antique	Kennard
person	discuss	everywhere	staircase	chart	gazard
feel	park	deny	daily	limp	fishlock
round	path	shot	essential	permission	cantileen
early	tower	refer	associate	headlong	gillen
table	weather	independent	conduct	violent	pardoe
question	wheel	feeling	relative	fade	frequid
effect	whole	bullet	upward	rake	hobrow
market	perform	juice	publish	trunk	candlin
woman	pity	nod	insult	mercy	litholect
stand	probable	gentle	cardboard	anxious	gumm
believe	signal	slip	humble	pedestrian	alden
fine	dish	diamond	contract	arrow	treadaway
instead	earn	press	mount	feeble	sumption
procedure	sweat	provide	tube	sorrow	horozone
group	trick	drum	moreover	brighten	hyslop
arrive	manage	reasonable	crisis	dam	mamomize
difficult	mud	boil	jug	outlet	horobin

Figure 3.4. Checklist format in the X-Lex Vocabulary test (from Milton, 2009: 254)

The X-Lex Test (above) and the Eurocentres Vocabulary Size Test (Meara & Buxton, 1987; Meara & Jones, 1987) are two widely known tests – alternative to the Vocabulary Levels Test – that aim to give an estimate of vocabulary size within specified frequency limits. The X-Lex tests 100 words from the first five 1000 word frequency bands, with 20 words randomly selected from each band. The test contains 20 made-up words, as shown in the figure above, to allow the score on the real words to be adjusted for guessing.²⁸

The latter of the two tests, i.e. the Eurocentres Vocabulary Size Test, is administered by computer rather than as a pen-and-paper test (the pen-and-paper

²⁸ The authors (Meara & Milton, 2003; Milton, 2009) recommend that the total vocabulary size be calculated in the following manner. First, the number of Yes responses to the real words is to be summed and multiplied by 50 to obtain a raw score out of 5,000. Second, the number of Yes responses to made-up words is to be added up and multiplied by 250. The final step involves adjusting the score on the real words by subtracting the score on the made-up words.

version is referred to as the Yes/No Test). It probes knowledge of around 150 words and gives an estimate of the test-takers' knowledge of the most frequent 10,000 words. Made-up words are also introduced in this test design in order to prevent an overestimation of the learners' knowledge. Its distinctive feature as compared with the X-Lex Test is that the test-takers are presented with each word one by one and have to indicate whether or not they think they are familiar with each word by pressing the Yes or No buttons. The X-Lex Test has the equivalent of the Yes button, i.e. putting a tick against a known word, but it does not offer the No option. Lack of a tick next to a word may or may not indicate that the test-taker does not know the word. Some words may, for example, be missed by mistake. The problem of omitted responses in the X-Lex Test can be remedied relatively easily, though, by replacing the checklist format with the Yes/No format. The Yes/No format explicitly rules out the possibility of non-responses (Meara cited in Beeckmans et al., 2001: 239), especially when the test is administered electronically.

Translation is another rather popular format in vocabulary breadth research. In tests of this format the test-takers are presented with an L2 target word and a set of translations or definitions in L1, and are expected to match the L2 word with the right translation or definition. A number of such bilingual versions exist, for example English-Korean, English-Mandarin and English-Vietnamese²⁹. The following extract comes from the English-Korean version of the 2007 Vocabulary Size Test (Nation & Beglar, 2007):

1. see: They **saw** it.
 - a) 잘랐다
 - b) 기다렸다
 - c) 보았다
 - d) 시작했다

2. time: They have a lot of **time**.
 - a) 돈
 - b) 음식
 - c) 시간
 - d) 친구들

²⁹ The bilingual versions in questions can be retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>.

To the best of our knowledge, an English-Polish version of the 2001 Vocabulary Levels Test or the 2007 Vocabulary Size Test has not yet been developed.

3.4.3. Productive measures of vocabulary breadth

It is widely acknowledged that the sizes of learners' receptive and productive vocabulary differ. Tests of productive vocabulary breadth are known to yield smaller figures than equivalently constructed tests of receptive vocabulary breadth (e.g. Fan, 2000; Milton, 2009: 117; Nation, 1990: 48; Waring, 1997). Measuring the receptive aspect of vocabulary exclusively is likely to result in overestimated figures of overall vocabulary breadth.

We turn now to an examination of formats used in productive tests of vocabulary breadth. Productive knowledge is not a simple Yes or No phenomenon. Learners may know a word but may be reluctant to use it as a result of uncertainty about its usage, pronunciation or spelling. They may know a less frequent word but may decide to use its more frequent synonym as a safer option. It is not surprising, therefore, that there are discrepancies, especially in relation to lower frequency bands, between the learners' productive vocabulary estimated on the basis of forced answer measures, such as fill-in-the-blank instruments, and estimations on the basis of samples of free writing. Meara (2005) argues that the former approach gives more accurate measurements of vocabulary breadth. Other researches, such as Laufer and Nation (1999), and Laufer (2005), hold the view that neither of the methods should be viewed as more accurate than the other, as they tap two different aspects of productive knowledge. They refer to the two different types of vocabulary knowledge as *free productive ability* and *controlled productive ability*, the latter being the focus of the thesis:

We refer to the ability to use a word at one's free will as free productive ability. (...) We use the term 'controlled productive ability' for the ability to use a word when compelled to do so by a teacher or researcher, whether in an unconstrained context such as a sentence-writing task, or in a constrained context such as a fill-in task where a sentence context is provided and the missing target word has to be supplied. (Laufer & Nation, 1999: 37)

Controlled productive ability, namely the ability to provide a target word when prompted to do so, is measured by translation and forced answer measures, such as the

fill-in-the-blank procedure. In a translation task the test-taker is presented with a list of L1 words and is expected to provide their L2 equivalents. Forced answer measures include C-tests and fill-in-the-blank tests. C-tests make use of authentic language materials varying in difficulty (e.g. Jafarpur, 1999; Klein-Braley & Raatz, 1984). When an authentic text is selected, the first and the last sentences are left intact. Starting from word two in sentence two, the second half of every other word is deleted. When the word has an odd number of letters, the number of letters provided is smaller than the number of letters deleted. Fill-in-the-blank tests make use of sentences rather than whole texts. A meaningful sentence context is presented for each target word, but the number of letters left after the deletion of part of the word is decided on the basis of the number of alternatives to the tested word.

The productive Vocabulary Levels Test (Laufer & Nation, 1999), a productive test of vocabulary breadth, makes use of the fill-in-the-blank procedure. The overall structure of the test was modelled on the receptive Vocabulary Levels Test. Like the receptive Levels Test, it contains a few frequency levels: 2,000, 3,000, 5,000 and 10,000-word frequency levels, and the University word level³⁰. Each frequency section elicits knowledge of 18 target words. The following sentences come from the 2,000 level:

The pirates buried the trea_____ on a desert island. (treasure)
Her beauty and cha_____ had a powerful effect on men. (charm)
La _____ of rain led to a shortage of water in the city. (lack)

The test was designed in relation to the same frequency lists as its receptive equivalent, namely Thorndike and Lorge's list (1944), the General Service List (West, 1953), Kučera & Francis (1967) and the University Word List (Xue & Nation, 1984). However, unlike the receptive Levels Test, the items were not sampled randomly from the above lists. The authors reveal that they used the lexical items from the original Levels Test and from the three parallel versions of the Levels Test made by Norbert Schmitt (Laufer & Nation, 1999: 37). This has resulted in a certain degree of overlap between items tested in the receptive and productive tests. In other words, not all items that are measured receptively are measured productively and vice versa (for more detail see Chapter 5, Section 5.3.3 and Appendix G, Tables A.4-A.6).

³⁰ Unlike the receptive Vocabulary Levels Test (Nation, 2001), the productive Levels Test (Laufer & Nation, 1999) does not contain a separate section that probes knowledge of words from the 1,000-word frequency band.

3.4.4. Advantages and disadvantages of selected formats of vocabulary breadth tests

Various existing formats used to measure vocabulary breadth have been outlined in the preceding paragraphs. This section presents a more critical view of the test selected for use in our experiment (the receptive and productive Vocabulary Levels Test) and other alternative measures testing the same constructs (translation-based tests of vocabulary size, Yes/No and checklist vocabulary tests). We are going to start off by weighing up the pros and cons of translation, as this method has possibly the longest history of the various vocabulary measures presented in this chapter.

Translation-based tests of vocabulary breadth have a number of advantages, one of them being time-efficiency. Their undisputed strength is that a relatively high number of words can be presented and marked in a short span of time. The test items are frequently presented in a decontextualised fashion (e.g. Kudo, 2001), which allows the test-takers to respond in a way that does not draw on L2 knowledge other than that directly related to the test items (Nation, 1998 cited in Kudo, 2001: 80). When context is provided, it is usually minimal (e.g. the English-Korean version of the 2007 Vocabulary Size Test, Section 3.4.2), which enables making inferences about the part of speech of the target word but not about the meaning itself. Additionally, when receptive and productive tests contain the same test items, receptive and productive knowledge of the target words can be directly compared. In the light of the above, translation may appear a very attractive option to SLA researchers and classroom teachers, especially where direct L1 and L2 equivalents exist.

Translation is not as straightforward a measure of vocabulary size, however, as it may at first appear. Problems arise when there is only partial equivalence between L1 and L2 words or when a word cannot be easily translated from one language into the other. If tests of vocabulary size were to be limited to direct L1 and L2 equivalents only, they would not test the construct they claim to tap. In a passive test format this can be partially circumvented by employing L1 definitions for those test items that do not have direct equivalents, see for example the English-Korean version of the 2007 Vocabulary Size Test (Nation & Beglar). Also, it has to be borne in mind that L2 learners do not necessarily have to translate every newly encountered L2 word back into their L1 to be able to grasp the meaning of the word. As L2 proficiency increases

learners are increasingly likely to pick up new words and phrases without having to constantly draw on their L1 mental lexicon. Moreover, depending on the balance between L1 and L2 use, L2 learners living in an L2-speaking country may, on occasions, have difficulty accessing relevant L1 equivalents (see the volume edited by Cook, 2003). For the above reasons, we felt that a translation-based test was not the best option for the subjects participating in the current study (adult Polish learners of English resident in Ireland). Consequently, we turned our attention to two other increasingly popular formats for measuring vocabulary size, namely Yes/No and checklist tests.

Yes/No and checklist tests, such as the X-Lex Vocabulary Test (Meara & Milton, 2003), the Yes/No Test (Meara & Buxton, 1987) and its computerised version, the Eurocentres Vocabulary Test, have the advantage of being far less time consuming than translation-based formats. However, their reliability is a matter of some concern. On the one hand, there is empirical evidence showing that scores sampled from individual sets of results of Yes/No tests using different selections of words produce strong correlations (Adamopoulou, 2000, quoted in Milton, 2009: 72). On the other hand, there are studies suggesting that the strength of such correlations may be inflated by the fact that some learners tend to overestimate their knowledge in self-reports, while others systematically underestimate it, producing a considerable test bias (Beeckmans et al., 2001; Eyckmans, 2004; Eyckmans, Velde & van Hout, 2007; Janssens, 1999). The reliability of Yes/No and checklist tests can vary depending on a range of factors, including *inter alia* learners' attitude towards the test (Eyckmans et al., 2007), learner cultural and linguistic background (Al-Hazemi, 1993; Meara & Buxton, 1987) or proficiency level (Read, 1997). Also, since the target words are presented in isolation, the test-takers are likely to occasionally mistake an unfamiliar word for a known one on the basis of a spelling resemblance. Anderson and Freebody's (1983) study, for example, demonstrates how some unknown words can be mistakenly interpreted as known words (e.g. "sham" for "shame"), which in turn results in inflated figures of vocabulary size. It appears, therefore, that students' self-reports as to which words they know should be looked at with caution (Nation, 1990).

The construct validity of Yes/No and checklist tests is another reason for concern. One way of establishing the construct validity of new language tests is to

correlate their results with scores obtained on other, preferably well-established, tests that are believed to measure the same construct. Discrepancies emerge between scores on individuals' self reports and scores obtained on other types of vocabulary size tests. The correlation between performance on the Yes/No test and the Vocabulary Levels Test, for example, has been shown to be non-significant (Cameron, 2002). The correlations between the scores on a translation test and various scoring methods for the Yes/No Test are also rather modest and range between .3 and .5, even between the sections of the tests that share the same content (Eyckmans, 2004). To put it into perspective, learners can fail to provide correct translations for as many as nearly 50 per cent of target items that they classified as known in the Yes/No Test (Eyckmans et al., 2007: 74).

One of the reasons for these discrepancies between Yes/No Test scores and scores on other tests of vocabulary size is that the Yes/No Test not only measures vocabulary breadth but also reflects the learners' self-assessment of their knowledge of vocabulary. The element of self-assessment that is part of the Yes/No format is claimed to be responsible for an observed response bias (Eyckmans et al., 2007). The discrepancies may also be partially brought about by the type of correction method used to control for the effects of guessing. The correction formulae that have been proposed in the literature for marking Yes/No tests make use of the number of Yes-responses to nonwords incorporated in the test. High rates of Yes-responses to nonwords are known to be problematic, as they lead to divergent scores when different correction models are employed³¹ (Huibregtse, Admiraal & Meara, 2002; Mochida & Harrington, 2006). Mochida and Harrington's (2006: 86) study is a very good illustration of this problem. A small mean score advantage of the Vocabulary Levels Test over the Yes/No Test is detected, but when two out of five correction methods investigated by the authors are applied to raw Yes/No Test results this difference turns out to be non-significant.

Having considered the pros and cons of translation, the Yes/No and checklist formats, we shall now move on to the receptive test format employed in our research project, namely the definition-matching format (the receptive Vocabulary Levels Test produced by Nation, 2001b and Schmitt et al., 2001). Unlike the Yes/No format

³¹ Another shortcoming that correction formulae suffer from is a marked drop in reliability that can result from the transformation of the original data (Beeckmans et al., 2001: 236).

definition-matching does not involve element of self-assessment, as it deals with actual language data rather than with students' perceptions of what they do or do not know. The problem of guessing, however, still remains. Tests, such as the receptive Vocabulary Levels Test, that require selecting one out of a few options may encourage guessing. According to Webb (2008), guessing is a considerable problem with the receptive Levels Test:

...there is a 17% chance that a subject will guess correctly on one response with no knowledge of any of the six choices, and the possibility of guessing correctly increases with knowledge of some of the words. (Webb, 2008: 81)

Milton (2009: 75) criticises the receptive Vocabulary Levels Test for not employing any correction formula to control for guesswork. He holds the view that the results from the receptive Vocabulary Levels Test owe more to informed guesswork than it would seem at first sight. A study by Schmitt et al. (2001) appears to suggest otherwise. A detailed comparison of post-interview scores and the actual test scores leads Schmitt et al. (*ibid.*: 67) to conclude that "guessing is not a serious problem with this format and that correct answers do reflect some underlying knowledge of the target word."

Comparable means and variances obtained during the pilot stage of the present study (described in Chapter 5) also indicate that test-takers obtain rather stable results regardless of the test version, which would be rather unlikely if a huge amount of guesswork were taking place. We are inclined to believe that in some instances test-takers may choose to make an informed guess, but that guessing does not change the scores significantly, provided of course that the test-takers take the test seriously (see e.g. Nation, 2007, for more on learner attitude and variability of scores). Two other potential problems with the Vocabulary Levels Test, i.e. the presence of a few Polish-English cognates and instances of drawing on general knowledge, are addressed in our account of the pilot study (Chapter 5, Section 5.3.3).

Devising relatively simple tasks that could produce the quantities of vocabulary necessary to make reasonable estimates of productive vocabulary breadth is considered something of a challenge. Assessing productive vocabulary appears more difficult than assessing receptive vocabulary because "the vocabulary produced by a learner, whether

in written or spoken form, tends to be so context-specific that it is difficult to calculate from any small sample the true size or range of the learner's productive vocabulary” (Meara & Fitzpatrick, 2000). The productive Vocabulary Levels Test, one of the principal methods of estimating productive vocabulary breadth, has been shown to be a reliable measure of productive vocabulary (Laufer & Nation, 1999) and to have good construct validity in the first 3,000 frequency bands (Fitzpatrick, 2007: 124-125). However, when it comes to the number of test items per level, the productive Vocabulary Levels Test elicits 18 words only from each frequency band, i.e. from the 2,000, 3,000, 5,000, University Word List and 10,000 bands. This has a positive and negative side to it. On the one hand, the test is fairly time-efficient. On the other, it can be argued that 18 items at each frequency level is a relatively small number. Ideally, we would prefer to have worked with a receptive and productive test that had the same number of test items.

3.5. Vocabulary breadth and the Common European Framework

The Common European Framework of Reference for Languages (CEFR) is part of our research design in that proficiency levels were defined in relation to the CEFR guidelines. The CEFR was designed by the Council of Europe in order to provide “a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe” (Council of Europe, 2001: 1). The development of CEFR was recommended in 1991 by the Rüschlikon Intergovernmental Symposium hosted by the Federal Swiss authorities so that “learners, teachers, course designers, examining bodies and educational administrators [could] co-ordinate their efforts” (Council of Europe, 2001: 6). The framework is believed to be a very useful tool in “communicating in rather concrete terms about what language learners can and cannot do” in a foreign language (Hulstijn, 2007: 663). Despite having some minor limitations (see e.g. volume 91 of the *Modern Language Journal*, 2007, or a recent Ph.D. thesis by Lin, 2009: Chapter 2), the CEFR has become a point of reference in L2 proficiency assessment across different languages (e.g. Little, 2007; Milton, 2006a). Many proficiency tests – including the one employed in our research project (OUP test, 2001) – provide guidelines on how obtained scores translate into the CEFR levels.

L2 proficiency within the CEFR is defined in terms of three broad bands, namely A (basic user), B (independent user) and C (proficient user). These three bands

correspond to the more traditional proficiency levels, i.e. beginners, intermediate and advanced respectively. The three bands are further subdivided into two levels each, which results in six bands in total, i.e. A1 (Breakthrough), A2 (Waystage), B1 (Threshold), B2 (Vantage), C1 (Effective Operational Proficiency) and C2 (Mastery). The CEFR guidelines contain information about the overall communicative proficiency at each level (see Table 3.3 below for A2 and B1 overall descriptors), and proficiency in terms of five communicative skills (i.e. listening, reading, spoken interaction, spoken production and writing). The CEFR descriptors are based on the notion of general competence (such as e.g. skills, know-how and ability to learn) and communicative language competence (e.g. linguistic, sociolinguistic and pragmatic). Different types of language skills (i.e. reception, production, interaction and mediation) and domains of language use (i.e. public, private, educational and occupational) are incorporated in the framework as well.

Table 3.3. The CEFR overall proficiency descriptors for the A2 and B1 levels

CEFR level	Description – global scale
A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

The notion of linking vocabulary breadth knowledge to CEFR proficiency levels received some attention at the stage when the CEFR descriptors were being formulated. The early CEFR descriptors included two word lists, one for the A2 level and the other for the B1 level. The lists in question are referred to as Waystage A2 list (van Ek & Trim, 1990b) and the Threshold B1 list (van Ek & Trim, 1990a). Waystage and Threshold lists were developed not only in English but also in a number of other languages such as German, French, Italian and Spanish. The Waystage lists across languages contain approximately 1,000 words, while the Threshold lists contain about 2,000 words on average. The Threshold lists exhibit more variation in terms of their size. The German and English Threshold lists contain the highest number of words, i.e. 2,400 and 2,200 respectively. The Italian and French lists have a slightly smaller number of words, namely 1,800. Spanish lists are significantly smaller containing only

about 800 words (Milton, 2009: 175). Waystage and Threshold lists have not been withdrawn, and yet they appear to have been somewhat pushed into the background in recent Council of Europe publications on the CEFR.

The Waystage and Threshold lists provide an indication of what kind of vocabulary L2 learners at different CEFR levels would be expected to know, but they do not specify vocabulary breadth figures for each proficiency level. In order to link vocabulary size figures to successive CEFR levels Meara and Milton (2003) analysed scores on their X-Lex test and compiled vocabulary breadth guides. Their figures suggest that considerable English vocabulary knowledge is needed to progress from the A1 level to the A2 level, namely about 1,500 words. After the initial 1,500 word gain, progressing with each subsequent CEFR level appears to involve smaller gains of about 500 words. The vocabulary guides that Meara and Milton (2003: 8) put forward appear to be in line with the data obtained in EFL (English as a Foreign Language) examinations in Greece and Hungary (see Table 3.4 below). The link between CEFR levels and vocabulary breadth, however, requires further investigation with the use of more varied testing procedures. The breadth test employed by Milton and colleagues (Meara & Milton, 2003; Milton, 2006a; Milton, 2006b; Milton, 2006c) in their attempt to capture students' vocabulary breadth at different proficiency levels was the X-Lex test – a testing procedure that involves a form of self-assessment of one's vocabulary knowledge (see Section 3.4.4 for a critical evaluation). It would, therefore, seem advisable to confirm the findings using testing procedures that do not entail an element of self-assessment.

Table 3.4. The CEFR levels and average EFL vocabulary scores

CEFR level	CEFR wordlist size (EFL)	Guides based on X-Lex (EFL)	Greece (EFL examinations)	Hungary (EFL examinations)
A1		< 1,500	1477	
A2	1,000	1,500 – 2,500	2156	
B1	2,200	2,500 – 3,250	3264	3136
B2		3,250 – 3,750	3305	3668
C1		3,750 – 4,500	3691	4340
C2		4,500 – 5,000	4068	

Despite the above limitation, the figures suggested by Meara and Milton (2003) constitute helpful guidelines for SLA researchers and language teachers. As mentioned earlier, the selection of frequency bands for our own project (described in Chapter 5, Section 5.3.3) was to some extent influenced by the size of CEFR word lists (Waystage and Threshold), and the research carried out by Milton and colleagues in relation to the

CEFR levels, vocabulary breadth and the plateau phenomenon within frequency bands (Section 3.3.3).

3.6. Phonological Short-term Memory (PSTM) and L2 vocabulary knowledge/learning

Having discussed the notion of vocabulary breadth and word frequency, we shall now focus on different methodologies adopted to analyse the relationship between PSTM and vocabulary learning. The methodologies used to explore this relationship involve studies of individual differences in PSTM capacity, experimentation, and case studies. Our own research project is a correlational study of individual differences. The studies presented later in this section that fall into the same category use a variety of vocabulary breadth measures, none of which, however, incorporates the aspect of word frequency. To the best of our knowledge, the present study is the only one that explores possible links between differences in PSTM capacity and gains in high-frequency vocabulary breadth.

This section addresses the need to further explore the causal direction of the association between PSTM and adult L2 vocabulary learning at elementary (A2) and pre-intermediate (B1) levels. The findings presented here draw mainly on L2 data, but some relevant L1 studies are examined also. The methodology used in one L1 study by Gathercole et al. (1992) is discussed in greater length order to explore the issue of the causality of the relationship between PSTM and vocabulary learning over time. We argue that the methodology used in the Gathercole et al. study, i.e. cross-lagged correlational analysis of longitudinal data, can be a useful tool for collecting further evidence on the direction of casualty in the adult population.

3.6.1. Early studies on PSTM capacity and vocabulary learning

The phonological loop (PSTM), typically tapped by tasks such as Nonword Repetition, Serial Nonword Recall, Serial Nonword Recognition, Word Span or Digit Span, has been linked to the pace of vocabulary acquisition and grammatical learning in L1 and L2. The evidence in this area draws on findings from different settings – laboratory (e.g. Papagno & Vallar, 1992; Service & Craik, 1993) and formal (e.g. Cheung, 1996; Service, 1992; Speciale et al., 2004) – as well as from various groups of

subjects – normally developing children, children with language impairments, normal adults and adults with brain injuries (e.g. Baddeley et al., 1998; Gathercole & Baddeley, 1990; Gupta, 2003; Jarrold et al., 2004; James Montgomery, 1995).

The first evidence of the association between PSTM and vocabulary learning comes from patient P.V. who suffered from a selective deficit in PSTM storage (Baddeley, Papagno & Vallar, 1988). Prior to this study long-term learning was usually tested via arbitrary sequences of familiar words, not phonologically unfamiliar material; such an approach presented ample opportunities for semantic coding (Baddeley et al., 1998: 163). Baddeley and colleagues, therefore, decided to conduct an experiment in which an unfamiliar language, Russian, was used. Their patient's capacity to learn new words, as indexed by an L1-Russian word paired associate task, was revealed to be seriously impaired. In contrast, the very same patient's performance on an L1-L1 word paired associate task remained entirely preserved. It appears then that P.V.'s PSTM deficit was associated with a specific impairment in long-term learning of phonologically unfamiliar words, while her long-term learning of novel combinations of familiar words was within the norm (also see Trojano, Stanzione & Grossi, 1992, for a description of patient S.C.). A similar pattern was identified in a graduate student, S.R., who had a very unusual STM profile (Baddeley, 1993). Despite being highly motivated and intelligent, he reported being unable to learn foreign languages. His performance on nonword repetition was very poor as compared to that of his colleagues, while his capacity for visual STM was quite normal. S.R. was tested using the paradigm developed for P.V., in which one task involved learning L1-L1 word pairs and the other memorising L1-L2 word pairs. Again a similar pattern of results was obtained. Despite his excellent use of mnemonics, S.R.'s performance on the latter task was seriously impaired when compared with control participants.

3.6.2. Language proficiency and strength of relationship between PSTM and vocabulary knowledge/learning

There is a body of evidence indicating that the involvement of PSTM in vocabulary knowledge/vocabulary learning diminishes with increased familiarity with L1 (Gathercole et al., 1992; Papagno & Vallar, 1995; Speciale et al., 2004: Experiment 1). There is also some evidence suggesting that a similar pattern may be observed in adult L2 learners (O'Brien et al., 2006), but the relationship between PSTM and L2

vocabulary knowledge/learning has not yet been explored at different L2 proficiency levels. Gathercole (2006a) argues that advanced L2 learners no longer need to be reliant on PSTM, as they can use their mental lexicons to access phonological representations of close neighbours of a new word:

...relatively experienced second language learners are able to use their substantial lexicons to mediate learning by access to lexical phonological representations of close neighbors, rather than relying on the more basic phonological learning mechanism tapped by nonword repetition. Such lexically supported learning has the advantage of capitalizing on knowledge of structures (which may be semantic, conceptual, or phonological in form) that have already been constructed. (Gathercole, 2006a: 515)

In relation to the L1 context, the link between PSTM measures and L1 vocabulary knowledge / learning has been shown to be significant in children (i.e. less proficient L1 users) but not in adults (i.e. more proficient L1 users). For example, in children aged 8 to 11 ($n = 45$) participating in a cross-sectional classroom-based study (Masoura & Gathercole, 1999) an L1-based (Greek) Nonword Repetition task correlated with L1 vocabulary scores at the .01 level ($r = .50$). In young adults ($n = 20$) involved in a cross-sectional study (Papagno & Vallar, 1995) Nonword Repetition did not correlate with L1 (Italian) low frequency vocabulary knowledge ($r = -.11$, $p > .05$). A laboratory-based study of adults ($n = 38$) by Speciale et al. (2004: Experiment 1) also indicated an absence of a significant link between a Nonword Repetition task and an L1 (English) low frequency vocabulary knowledge task ($r = .15$, $p > .05$). Additionally, a longitudinal study of children ($n = 80$) by Gathercole et al. (1992) demonstrated that the link between L1-based (English) Nonword Repetition and L1 vocabulary knowledge gradually weakens as children grow older (e.g. $r = .419$, $p < .01$ and $r = .284$, $p < .05$ for ages 5 and 8 respectively).

In relation to child L2 learning, the relationship between PSTM and child L2 vocabulary knowledge/learning also appears to be stronger at a lower stage of L2 development as compared to higher stages. A longitudinal study by Service (1992) involving Finnish primary school children ($n = 44$) examined initial stages of L2 (English) learning in a classroom setting and revealed that normal variation in L2-based

Nonword Repetition³² was closely related to the students' final grades ($r = .66$, $p < .001$). Nonword Repetition was also shown to highly correlate with a number of subskills in L2 (all $ps < .001$). However, neither L2 vocabulary knowledge nor L2 vocabulary learning was tested directly in the Service study. Another study (Cheung, 1996) addressing the issue of a link between PSTM (L2-based Nonword Span) and L2 (English) learning (tapped via a paired-associate learning task) in Chinese primary school children³³ ($n = 84$) discovered a significant link between PSTM and the number of learning trials of novel L2 vocabulary ($r = -.45$, $p < .001$). Interestingly, when the students were divided into a low L2 vocabulary subgroup and a high L2 vocabulary subgroup the relationship remained significant for the low-vocabulary subgroup only. The observed pattern could be interpreted as an indication that the weakening of the relationship between PSTM and L1 vocabulary knowledge as the L1 proficiency increases reported by Gathercole et al. (1992) may also be applicable to the L2 context (but see Masoura & Gathercole, 2005). It should be pointed out, however, that while the Gathercole et al. study examined the pattern of correlations between relevant variables at different levels of L1 proficiency (between ages 4 and 8), in the Cheung study the subjects' L2 proficiency level was not measured directly.

In relation to adult L2 learning, laboratory-based studies involving adults indicate that there is a strong link between PSTM (mostly tapped by Nonword Repetition) and L2 vocabulary learning at the onset of L2 learning (Craik & Watkins, 1973; Gupta, 2003; Papagno & Vallar, 1992; Speciale et al., 2004: Experiment 1; see also the older adult part of the Service & Craik study, 1993). As for classroom-based longitudinal studies on adult L2 learning (O'Brien et al., 2006; Speciale et al., 2004: Experiment 2), they present a rather mixed picture. O'Brien et al. demonstrate that L1-based Serial Nonword Recognition scores at Time 1 correlate with free productive knowledge of L2 vocabulary at Time 2 in lower ability learners, but not in higher ability learners ($r = .423$, $p < .05$ and $r = .175$, $p > .05$ respectively)³⁴. Speciale et al., on the other hand, detect no significant link between a Nonword Repetition task (based neither on L1 nor L2) at Time 1 and receptive L2 vocabulary knowledge at Time 2 in

³² L2-based Nonword Repetition was measured four times at one-year intervals and then averaged. L1-based Nonword Repetition scores were also recorded. However, the variation in these scores turned out to be insufficient, which led the author to exclude the scores from her analyses.

³³ The average age of the subjects was 12.2, and they had been learning L2 in the school context since the age of 3 or 4.

³⁴ The exact L2 proficiency level was not reported. The learners were divided into lower and higher ability groups according to the results they obtained on an entry L2 test (the SAT II Spanish test).

adults at the A1 level of L2 proficiency. The difference between the results obtained by O'Brien et al. and those obtained by Speciale et al. may have been brought about by the nature of the L2 vocabulary measures used. O'Brien et al. measured free productive knowledge of L2 vocabulary in spoken output, while Speciale et al. measured the students' ability to arrange some high frequency words into 10 pairs that could be reasonably contrasted (Speciale et al. do not provide examples). As only certain words can be contrasted in terms of their meaning, it is not unlikely that the Speciale et al. receptive L2 vocabulary task was too restricted in its content to reflect the amount of progress their students made over a 7-week period of L2 instruction.

Neither O'Brien et al. (2006) nor Speciale et al. (2004: Experiment 2) deal with the link between PSTM and different levels of L2 proficiency in adult learners. If the pattern observed by O'Brien et al. among their high and low ability adult subjects is replicable in adult A2 and B1 learners, then in the current study we would expect to discover a stronger relationship between L1-based PSTM measures and L2 vocabulary learning at the A2 level and a weaker (or no) relationship at the B1 level. If, however, the pattern observed by Speciale et al. in their adult A1 learners is replicable, then we would expect no relationship between L1-based PSTM measures and L2 vocabulary learning at the A2 level or at the B1 level.

3.6.3. The causal direction of relationship between PSTM and vocabulary knowledge/learning

Many PSTM studies on adult subjects are correlational in nature. As leading researchers in the field admit (see e.g. Baddeley et al., 1998: 160), correlation can indicate an association between two factors, but it does not imply causation. Even though PSTM is generally regarded as a strong *predictor* of word learning (e.g. Baddeley, 2003b; Papagno et al., 1991; Service, 1992), the direction of causation in this well-established link is still open to debate.

In principle it is plausible that good PSTM functioning supports vocabulary development or that the reverse holds true. If the link runs in the opposite direction to that suggested by Baddeley and colleagues, then it is word learning that drives PSTM performance rather than vice versa. One of the arguments put forward in favour of the view that individual differences in linguistic knowledge drive PSTM performance is

that PSTM measures are known to be influenced by LTM (Chapter 2, Sections 2.3.2 and 2.3.3). The effects of ‘lexicality’ (better recall for words than nonwords) and ‘wordlikeness’ (better recall for nonwords resembling real words) indicate that at a recall stage serial memory for items is affected by top-down redintegrative processes drawing on long-term lexical knowledge (Gathercole et al., 2001). Bearing the above in mind, individuals with lower vocabulary skills might be expected to obtain lower PSTM scores as compared to those with better developed lexicons, while adults with a richer knowledge of languages (L1 and others) might be expected to have more efficient PSTM. In Papagno & Vallar’s cross-sectional study of polyglots (1995), for example, polyglot Italian speakers were shown to outperform their non-polyglot control group on both Digit Span and Nonword Repetition, but no significant differences were detected on visuospatial memory measures and nonverbal ability measures. According to Baddeley and colleagues this study implies that “a natural talent for language learning may arise directly as a consequence of excellent phonological loop function” (Baddeley et al., 1998: 166), but this case does not in fact reveal whether good vocabulary knowledge supports PSTM or whether the reverse holds true.

There are two alternative accounts, the PSTM account and the lexical account³⁵, that propose opposite directions of causality in the relationship between PSTM and vocabulary knowledge/learning. The PSTM account (Baddeley et al., 1998) posits that PSTM is a causal determinant of both Nonword Recall/Recognition ability and vocabulary knowledge/learning. Within this account a correlation between Nonword Recall/Recognition ability and vocabulary knowledge/learning is believed to reflect the effect of the underlying variable, i.e. PSTM capacity. The PSTM account is based on the premise that learning a new word involves the same processes that encompass those employed in nonword tasks. After all, a new word is a nonword to a specific learner (Gathercole, 2006a; Gupta & Tisdale, 2009). While proposing that PSTM causally determines vocabulary learning, proponents of this view also acknowledge the impact of existing lexical/phonological knowledge on performance on nonword tasks (e.g. Baddeley, 2007; Gathercole, 1995; Gathercole et al., 1991). As mentioned earlier, repetition accuracy of nonwords is known to be influenced, *inter alia*, by their wordlikeness, i.e. the degree to which they resemble real words in a given language.

³⁵ The terminology is adopted from Gupta and Tisdale (2009).

The opposing account, called the linguistic account (Snowling et al., 1991), proposes that it is vocabulary size that causally determines Nonword Recall/Recognition ability. Within this account the focus is on factors other than PSTM that influence performance on nonword tasks, namely the extent of lexical/phonological knowledge that an individual possesses. Snowling et al. (1991: 372) argues that good L1 vocabulary knowledge allows children to better cope with the processing demands of nonword repetition tasks. As Edwards et al. (2004: 434) point out, "...an increase in vocabulary size does not simply mean that the child knows more words, but also that the child is able to make more and more robust phonological generalizations." Such phonological generalizations would be likely to boost performance on nonword tasks. Both accounts, therefore, acknowledge that the level of lexical/phonological knowledge may affect performance on nonword tasks. The major difference between the two accounts is that the PSTM account emphasises the causal role of PSTM in the growth of vocabulary knowledge, while the linguistic account posits that the involvement of PSTM may be relatively minor.

The methodologies used to address the issue of the causal direction of the connection between PSTM and adult foreign word learning include experimentation, studies of individual differences and case studies. With regard to experimental word learning a number of lab-based studies have been carried out, in which PSTM functioning was experimentally manipulated via different learning conditions, while word learning ability was tapped by an L1-L2 word paired-associate learning procedure (e.g. Ellis & Beaton, 1993; Ellis & Sinclair, 1996; Papagno et al., 1991). Experimental manipulation of PSTM efficiency involves articulatory suppression and articulatory rehearsal (Chapter 2, Section 2.3.2). Let us recall that articulatory suppression, that is engaging in irrelevant articulation during the presentation of memory items, is thought to diminish the effective capacity of PSTM (Murray, 1967), while articulatory subvocal rehearsal is believed to enhance PSTM by reinforcing temporary phonological representations (Baddeley, 1992: 8) and providing multiple redundant representations (Ellis & Beaton, 1993: 553). Interfering with PSTM functioning showed that adult foreign vocabulary learning in laboratory conditions is significantly enhanced by rehearsal and disrupted by articulatory suppression. The evidence that experimental manipulations of one factor (PSTM efficiency) bring about a significant change in the other variable (vocabulary learning) can be interpreted as supporting the view that

PSTM capacity is a predictor of word learning rather than its consequence at least at the initial stages of vocabulary learning.

Further evidence in favour of the PSTM account comes from a series of experiments concerned with the effect of phonological similarity (Chapter 2, Section 2.3.2) among memory items (Conrad & Hull, 1964) and memory item length (Baddeley et al., 1975) on adult foreign word learning. The phonological similarity effect constitutes an important argument for positing the existence of the phonological loop, and it has been used to put forward a theoretical proposal that PSTM employs a phonological code. The phonological similarity effect is a phenomenon observed during immediate serial recall of verbal material, namely that recall for lists of phonologically similar items is poorer than that for lists of phonologically dissimilar items. Assuming that the learning / memorisation of unfamiliar foreign vocabulary – as opposed to that of pairs of familiar L1 words – depends on the phonological loop, an increase in the level of phonological similarity among the items to be memorised should make phonologically similar words confusable and hence more difficult to recall. The assumption is that if new vocabulary learning was not dependent on PSTM capacity no such effect would be expected to occur. By manipulating the level of phonological similarity in a series of experiments involving adults Papagno and Vallar (1992) demonstrated that the phonological loop system (PSTM) is not involved in the learning of known L1 vocabulary pairs, but it is crucial in the learning of unfamiliar-sounding words from a foreign language. Increased phonological similarity among memory items did not impair the learning of familiar (L1) vocabulary pairs, but it did affect unfamiliar (L2) vocabulary learning.

Further attempts to provide support for the PSTM account come from studies using methodology based on individual differences in PSTM. Within this approach interference techniques such as e.g. articulatory suppression were replaced by repetition, recall and recognition tasks measuring the *actual* variation in PSTM efficiency (Chapter 2, Section 2.6). Three types of research projects were carried out within this methodology, and these include: a) lab experiments (e.g. Atkins & Baddeley, 1998; Gupta, 2003), b) longitudinal studies (O'Brien et al., 2006) and c) a combination or both (e.g. Speciale et al., 2004). The timeframe of longitudinal studies extended over a considerably longer period than that of the longest word learning lab

experiments. While there are quite a few longitudinal studies conducted on children (e.g. French, 2003; Gathercole et al., 1997; Masoura & Gathercole, 1999; Service, 1992), there is only a handful of such studies carried out among adults (O'Brien et al., 2006; Speciale et al., 2004: Experiment 2). Even though the O'Brien et al. study documented the predictive power of PSTM (at Time 1) on subsequent free productive use of L2 vocabulary (at Time 2) in adult learners, this kind of design on its own does not provide watertight evidence for causation (Field, 2005), and should probably be accompanied by a more rigorous approach, such as that adopted by Gathercole et al. (1992) and French (2003).

Gathercole and colleagues' (1992) study of young children learning English as an L1 extended over a period of 4 years, and explored possible mutual influences in both directions between PSTM and vocabulary learning. A cohort of 80 four-year olds from a middle class background was subjected to PSTM testing and receptive vocabulary testing. The tests were re-administered when the individuals reached the ages of 5, 6 and 8. To explore the directionality of the association Gathercole and colleagues applied a cross-lagged partial correlational analysis, which involved calculating and comparing the strength of the correlations between early PSTM and later receptive vocabulary, and between early vocabulary and later PSTM scores. According to the assumptions of a cross-lagged correlational analysis, the correlation should be stronger in the causal than in the noncausal direction (e.g. Campbell, 1988; Campbell & Stanley, 1963; Crano et al., 1972; Crano & Mellon, 1978). The pattern obtained for the link between PSTM and vocabulary at ages 4 and 5 suggests that PSTM capacity drives vocabulary learning rather than vice versa. However, this pattern shifted for age range 6 – 8 suggesting that expanding vocabulary knowledge might drive PSTM performance at higher levels of familiarity with the language.

Correlations between PSTM and L2 vocabulary learning in normal adults have not yet been compared in two causal directions across a significant time period. We argue here that the methodology adopted in L1 research by Gathercole and colleagues (1992) can help to determine whether PSTM performance is causally related to L2 vocabulary learning in adults at different levels of L2 proficiency. There are some reasons why the pattern of relationship between PSTM and L2 vocabulary in adults might differ from that observed in children in the study by Gathercole et al. First,

variation in PSTM scores could be expected to be smaller in adults as compared to children, which can render the link between these two factors non-significant in young adults (Service & Craik, 1993). Second, when PSTM measures are based on the subjects' L1, differences in L1 long-term knowledge in the two populations of subjects may result in different patterns of results. Third, unlike the subjects from the Gathercole et al. study, our multilingual adults, such as those in the current study, have more than one linguistic system at their disposal (see Section 3.2.3 for a discussion of the bilingual mental lexicon). Additionally, adults may use different strategies in Serial Nonword Recall and Serial Nonword Recognition than those observed in children.

Chapter 4: Collocations and Memory in Adult Second Language Learners

4.1. Introduction

The collocational dimension of the lexicon has been increasingly recognized; leading figures in bringing about such recognition include Firth (e.g. 1957 or 1974) and Sinclair (e.g. 1991 or 1996). In relation to language acquisition also collocational knowledge has increasingly been seen as playing a vital role – in respect of both first language (L1) acquisition (e.g. Ellis, 2001) and second language (L2) acquisition (e.g. Durrant, 2008). Moreover, collocational considerations have in recent decades loomed larger and larger in language pedagogy (e.g. Lewis, 2000; Nattinger & DeCarrico, 1992). This chapter locates the current study on the link between PSTM and collocational learning within a comprehensive account of theoretical linguistic work on collocational knowledge. The present chapter consists of five sections, four of which look at the notion of collocation in language learning, and one of which analyses some indirect evidence pointing to the fact that PSTM may play a role in L2 collocational learning in the classroom in a longitudinal perspective.

Section 4.2 starts off by presenting and analysing the Oxford English Dictionary definition of collocation. It describes Firth's notion of *meaning by collocation* and Sinclair's concept of *semantic prosody*. It also outlines two widely known principles of interpretation by Sinclair, namely the *idiom principle* and the *open choice principle*. Section 4.3 presents terminology used in collocation and formulaic language research. It elaborates on functions of collocations and different frameworks of classification. It contains an overview of two major approaches to collocations, namely the frequency-based approach and the phraseological approach. Section 4.4 discusses the theoretical relevance of collocations in language learning, and reviews models of L1 and L2 collocational learning in the context of formula-based approaches to language learning. Section 4.5 presents test formats used to tap L2 collocation knowledge and statistical procedures used to calculate the *frequency* and *strength* of collocations.

The relationship between PSTM and vocabulary learning in child L1 acquisition and adult L2 experimental learning has been discussed in Chapter 3, Section 3.6. The final section of this chapter, i.e. Section 4.6, presents some laboratory data (Ellis & Sinclair, 1996) linking PSTM with L2 collocation learning at the initial stages of adult L2 learning. The Ellis and Sinclair laboratory-based data, as well as some indirect classroom-based evidence (O'Brien et al., 2006; Speciale et al., 2004), are presented to support the view that PSTM may play a role in the development of collocational knowledge in adult L2 learners at the elementary level of proficiency (A2) in the classroom setting as well. Two different viewpoints are presented on the question whether or not PSTM is implicated in adult L2 collocation learning in the classroom setting at a post-beginner level, and more specifically at a pre-intermediate (B1) level. Some indirect evidence is presented to support the two points of view.

4.2. The notion of collocation

One initial indication of importance of collocations for a theory of lexicon is the fact that the majority of sentences appear to be “made up of interlocking collocations such that they could be said to reproduce, albeit with important variations, stretches of earlier sentences” (Hoey, 2005: 5). Despite the fact that the importance of the phenomenon of collocation is widely recognised, definitions of collocation have taken various shapes and forms. This section begins by outlining the difference between broad and narrow definitions of collocations provided in the Oxford English Dictionary. It presents Firth’s concept of *meaning by collocation* and Partington’s distinction between textual, associative and statistical definitions of collocations. Sinclair’s notions of *semantic prosody* and *principles of interpretation* are also dealt with in this section.

4.2.1. Defining and identifying collocations

The third online edition of Oxford English Dictionary (2008) defines the term *collocation* in its broad sense as “the action of setting in a place or position, *esp.* of placing together with, or side by side with, something else; (...) [f]requently applied to the arrangement of words in a sentence, of sounds, etc.” Along with this definition the Oxford English Dictionary provides a few examples of early references to the word collocation, three of which are listed below:

1750 HARRIS Hermes II. iv. Wks. (1841) 197 The accusative..in modern languages..being subsequent to its verb, in the collocation of the words.

1751 JOHNSON Rambler No. 88 p5 The difference of harmony arising..from the collocation of vowels and consonants.

1873 EARLE Philol. Eng. Tongue (ed. 2) p630 All languages use greater freedom of collocation in poetry than in prose.

From a linguistic perspective the above-mentioned definition is too vague to be of any practical application. A far more specific definition is required to represent the concept of collocation adequately.

Apart from the broad definition of collocation the Oxford English Dictionary also provides a technical definition of the concept that highlights three aspects of the term that the previous broader definition does not address. The technical definition refers to collocation as “the habitual juxtaposition or association, in the sentences of a language, of a particular word with other particular words; a group of words so associated” (2008). The first of the three aspects that it addresses is that only the juxtaposition of *certain* words, as opposed to that of *any* lexical items, can count as collocation. According to this line of reasoning, for example, “reach a conclusion” and “compare results” does not fall into the same category of collocability. The former would be viewed as collocation while the latter as a relatively free verb-noun combination (Howarth, 1998: 35 & 38). The second aspect this definition encompasses is that co-occurrence of words has to be ‘habitual’ for the constituent words to be classified as collocationally linked with one another. One-off or low frequency juxtapositions do not comply with this definition. Interestingly, this runs against the third Oxford English Dictionary quotation presented above (from 1873), which seems to suggest that that the link between rare word-pairings used in poetry can also be classified as collocation. The third aspect of the Oxford English Dictionary definition is that both the *connection* between particular lexical items and a *group* of such particular items can be referred to as collocation. Using “pay attention” as an example, the juxtaposition of “pay” and “attention” can be viewed as a product of collocation or the whole phrase can be called a collocation.

The use of the term *collocation* in linguistics is usually attributed to the British linguist John Rupert Firth³⁶. He introduced the concept of *meaning by collocation* in his often-quoted article “Modes of Meaning” (1957) by saying “I propose to bring forward as a technical term, meaning by ‘collocation’, and to apply the test of collocability” (ibid.: 194). The Firthian theory of meaning was developed by Firth in collaboration with anthropologist Malinowski (cf. Lyons, 1977: 607) and then further refined by his numerous followers. As the above quote indicates, the Firthian notion of meaning is more concerned with the contextual aspect than it is with “the conceptual or idea approach to the meaning of words” (Firth, 1957: 196). His very famous slogan “You shall know a word by the company it keeps” highlights the fact that Firth conceived of meaning as “a complex of contextual relations” (ibid.: 19), with the components of an utterance blended with each other to the extent to which they were no longer recognizable as separate units. According to Firth, the collocation within which a given word occurs constitutes part of the meaning of that word, as the following example illustrates: “one of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night*” (ibid.: 196).

According to Partington (1998), definitions of collocation fall into three categories, namely textual, associative and statistical. For example, Sinclair’s definition of collocation – i.e. “the occurrence of two or more words within a short space of each other in a text” (1991: 170) – can be referred to as *textual*, because it captures the linear aspect of language, but fails to distinguish between single instances of co-occurrence and habitual juxtapositions of lexical items. Leech’s (1974) definition of collocation – i.e. “the associations a word acquires on account of the meanings of words which tend to occur in its environment” (1974: 20) – can be classified as *psychological* or *associative*, as it reflects the Firthian syntagmatic aspect of meaning. In the psychological realm native speakers throughout their lives develop intuitions as to which collocations are more and which are less likely to occur in certain situations (Partington, 1998: 15-16). As for the *statistical* reality, when one lexical item is identified in a text, the other one can be expected to appear there as well “with greater

³⁶ It is Firth who is often credited with coining the term *collocation*, but it should be noted that the concept was used in linguistic discourse prior to Firth’s 1957 paper. The 1928 edition of *Webster’s New International Dictionary*, for example, contains a much earlier definition of collocation, which depicts collocation as “an arrangement or ordering of objects (esp. words) with reference to each other”. Palmer (1933 & 1934) also talks about the phenomenon of collocation in his report for the IRET (the Institute for Research in English Teaching).

than random probability” (Hoey, 1991: 7). This brings us to the notion of mutual expectancy of words (Firth, 1957: 195), according to which words that collocate with one another predict the occurrence of one another in a given sample of language. This frequency-based aspect of lexical co-occurrence falls into Partington’s statistical formulation of the notion of collocation (1998: 16).

4.2.2. Semantic prosody and near-synonymy

The importance of collocations in linguistics is partly derived from the fact that favourable and unfavourable connotations are quite often contained not in a single word but in the collocates³⁷ that word has. As Sinclair remarks, “many uses of words and phrases show a tendency to occur in a certain semantic environment” and some words like “happen” often occur with bad things such as e.g. accidents (Sinclair, 1991: 112). Partington gives two more examples to illustrate this phenomenon, the verb “commit” and the adjective “rife” (Partington, 1998: 66-67). Both words turn out to collocate with items of rather an unpleasant nature. The word “commit” co-occurs with negatively-charged items such as “deception”, “offences”, “serious crime” or “suicide”. The word “rife” co-occurs with “speculation”, “mistakes” or “crime”, etc.

The existence of an ‘intuitive’ meaning of an item that is manifested when the item is in context is labelled *semantic prosody* (Louw, 1993: 172). The term *semantic prosody* was first coined by Sinclair (1987b) in his publication on the COBUILD project (see Section 4.3.2), later adopted by Louw (1993), and then reformulated by Sinclair (1996) as an element of his model of extended units of meaning. In Louw’s formulation, certain features of a single lexical item are present in its surroundings, which limits the number of choices available to a language user. For instance, words that usually keep ‘bad company’ like “set in”, which is habitually associated with negative events (Sinclair, 1987b: 155-156), are not normally expected to occur in a favourable environment. The word “set in”, in fact, has rather few collocates that are neutral, while most are highly undesirable and unattractive, e.g. “ill-will”, “disillusionment”, “mannerism” or “slump”. The power of word groupings, like the ones specified in the preceding paragraph, is such that a word that has a certain affective meaning when surrounded by its typical collocates may retain the same

³⁷ Collocates are the words the lexical item under scrutiny co-occurs with. The term was coined by Halliday (see Section 4.2.3).

affective meaning when used with atypical collocates (Xiao & McEnery, 2006). This, in turn, enables the language user to achieve the effect of irony or humour by consciously violating the semantic prosody condition. Thus, “[a] phrase like *good times set in*” can be seen as “a highly marked, probably humorous use” (Partington, 1998: 67) owing to the violation of the semantic prosody condition of the phrase “set in”.

It is the collocational context that allows the language user to distinguish between different senses of the same word. As Hoey puts it (2005: 81), “the patterns of one use of a polysemous word always distinguish it from those of the other uses of the word.” Thus, the word “commit” when analysed in isolation could have an array of different meanings. When in context, as in “commit to memory”, “commit a crime” or “commit oneself”, the ambiguity is resolved and the meaning becomes apparent (Bartsch, 2004: 72). In similar vein, words with almost identical meanings can be distinguished from one another on the basis of their collocational profiles. Near-synonyms such as “strong” and “powerful” can be distinguished from one another by reference to their collocates – the former co-occurring with “tea”, the latter with “engine”, but not vice versa (Halliday, 1966). This applies to many other lexical items. Partington (1998: Chapter 2), for example, illustrates how the collocational networks of “sheer”, “pure”, “complete” and “absolute” can differentiate the four adjectives from one another. Finally, in addition to different collocational behaviour, near-synonyms can also differ in the type of prosodies they have; e.g. “flexible” is positive, while “fickle” is distinctly negative (cf. Xiao & McEnery, 2006: 108). Other pairs of words that could be quoted in this connection are e.g. “slim” / “skinny”, “elderly” / “senile” or “firm” / “pig-headed”.

4.2.3. Principles of interpretation and economy of effort

The idiom principle and the open choice principle

The widespread existence of collocations in various languages suggests that a proficient speaker of a given language – apart from adhering to grammatical rules – must have some level of mastery of multi-word units (Conklin & Schmitt, 2007: 4). The way lexical units are combined into larger chunks seems, as argued above, to be dependent on more than just grammatical acceptability. To account for this widespread existence of lexical chunks, and for the divergent collocational behaviour of

synonymous words, Sinclair (1991: 109-115) posits that the structure of language is influenced by what he terms the 'idiom' principle much more than it is by the 'open choice' principle.

The *open choice* principle is a more traditional linguistic approach to viewing language, according to which "when one unit is completed (a word or a phrase or a clause), a large choice opens up and the only restraint is grammaticalness" (ibid.: 109). If human language is a series of open-ended choices with grammaticalness being the only constraint, any word from a certain class of words should fit into a given slot. However, in reality, the number of options that can occur in one slot can be severely limited at times. Sinclair's *idiom* principle highlights the fact that there are restrictions with regard to the number of options to choose from when one unit is in place, and in some cases there is only one option available, even though a given multi-word phrase "might appear to be analyzable into segments" (ibid.). The degree of availability of open choice is known to vary from one point to another along the progression of an utterance. According to Sinclair discourse is created on the basis of the interaction between the idiom principle and the open choice principle, but it is the idiom principle that is the default mode of speech interpretation and production (1987a: 320-324).

Sinclair referred to the idiom principle and the open choice principle as "principles of interpretation" (ibid.: 319), a term which can itself be interpreted in more than one way. The principles could be seen as a set of directions that can help linguists explore the structure of a text. Alternatively, they could be viewed as a depiction of "how language is normally processed in the mind" (Durrant, 2008: 8). The quotation presented below represents the latter perspective, namely the processes in the mind of a listener as he/she attempts to decode an incoming message:

...the first mode to be applied is the idiom principle since most of the text will be interpretable by this principle. Whenever there is good reason, the interpretative process switches to the open choice principle, and quickly back again. Lexical choices which are unexpected in the environment will probably occasion a switch. (Sinclair, 1987a: 324)

The economy of effort principle

The application of the idiom principle by language users in every day communication is thought to arise out of “a natural tendency to economy of effort” (Sinclair, 1987a: 320). As Code states, “it would be physiologically impossible for us to produce speech with the rapidity and proficiency that we are able to if we had to plan and perform each segment individually” (1994: 139). The principle of economy gives us access to prefabricated chunks of lexis, so that we do not have to go through the labour of generating all the phrases from scratch. In that way language users can “direct their attention to the larger structure of the discourse, rather than [keep] it focused narrowly on individual words as they are produced” (Nattinger & DeCarrico, 1992: 32). This issue is closely related to a point made by Partington (1998) concerning the possible existence of two mental lexicons, one of single lexical items and the other of semi-preconstructed phrases. At first sight, it would seem counterproductive to store lexical information in that particular way, assuming of course that that is the way in which multi-lexical chunks are stored (see e.g. Underwood et al., 2004, for some evidence suggesting that chunks are stored holistically). On closer inspection though, as Partington argues, the reason behind this ‘double’ storage “must be that the effort saved by making far fewer slot-filling choices in *real-time* (...) easily outweighs the disadvantages of unwieldy storage” (Partington, 1998: 19; original italics).

Apart from their role in alleviating processing pressures in speech production, another powerful reason for committing multi-lexical chunks to memory is their contribution to language comprehension. In real-time language decoding message recipients tend to match the sound signal with their expectations instead of attending to each individual morpheme (Rost, 1990). When experiencing difficulty in decoding a message owing to e.g. noisy conditions or non-targetlike pronunciation, listeners also tend to draw on their lexicon of collocational relationships in order to fill in the gaps (Bonk, 2001: 116). Making use of conventional word pairings and fixed phrases in real-time decoding appears to cut down processing effort on the part of the listeners. Conventional word pairings provide the listeners with a certain framework of reference, and allow them to focus on other, more complex aspects of language processing. For example, analyses of the structure of TV weather forecasts and its impact on processing load (Hickey & Kuiper, 2000; Wray, 2002: 79-83) indicate that without such a

structural framework message decoding would require far more processing power on the part of the hearer.

The principle of economy of effort is often linked with the concept of nativelike selection. *Nativelike selection* implies that native speakers are able to produce and recognize phrases that are appropriate in certain contexts or registers, and that they “do not exercise the creative potential of syntactic rules to anything like their full extent (...) indeed, if they did so they would not be accepted as exhibiting nativelike control of the language” (Pawley & Syder, 1983: 193). As Pawley and Syder point out, non-native speakers who have not had much exposure to the L2 as it is used in the speech community are likely to produce sentences that are strictly grammatical but sound unnatural and rather unidiomatic (ibid.: 195). This is where collocational competence comes into play. As Pawley and Syder posit, one of the factors that creates the impression of nativelike control and fluency is the use of what they call *memorized sequences*:

Some clauses are entirely familiar, memorized sequences. These are strings which the speaker or hearer is capable of consciously assembling or analysing, but which on most occasions of use are recalled as whole or as automatically chained strings. (Pawley & Syder, 1983: 205)

Pawley and Syder admit, however, that despite the fact that in conversational talk memorized combinations are very frequent indeed, there are also many other types of clauses that fall somewhere between the two extremes, memorized sequences at one end and relatively free combinations at the other.

4.3. Collocations and formulaic language: terminology, approaches and functions

This section examines three important aspect of the concept of co-occurrence of words. First, it presents a number of terms used over the years to refer to different types of links between words, and contrasts a broader notion of formulaic language with a narrower notion of collocation. Second, it discusses two approaches to studying and classifying collocations, namely the frequency-based approach and the phraseological approach. Third, it highlights the significance of the co-occurrence of words in written

and spoken language by outlining a number of functions that collocations and formulaic language are believed to have.

4.3.1. Collocations and formulaic language

Research on collocations has been extremely diverse in terms of its methods and theoretical assumptions. “The morass of overlapping terminology used by various researchers over the years” (Bonk, 2001: 113) in the same or related areas of investigation has been detrimental to the emergence of a more general picture and to the reconciliation of the full range of findings within a single theoretical account (Wray, 2002: 5). Figure 3.1 contains a compilation by Wray of terms used to describe various aspects of the collocational links between words:

amalgams – automatic – chunks – clichés – co-ordinate constructions –
 collocations – complex lexemes – composites – conventionalized forms – F[ixed]
 E[xpressions] including I[dioms] – fixed expressions – formulaic language –
 formulaic speech – formulas/formulae – fossilized forms – frozen metaphors –
 frozen phrases – gambits – gestalt – holistic – hollophrases – idiomatic – idioms
 – irregular – lexical simples – lexical(ized) phrases – lexicalized sentence stems
 – listemes – multiword items/units – multiword lexical phenomena –
 noncompositional – noncomputational – nonproductive – nonpropositional –
 petrifications – phrasemes – praxons – preassembled speech – precoded
 conventionalized routines – prefabricated routines and patterns – ready-made
 expressions – ready-made utterances – recurring utterances – rote – routine
 formulae – schemata – semipreconstructed phrases that constitute single choices
 – sentence builders – set phrases – stable and familiar expressions with
 specialized subsentences – stereotyped phrases – stereotypes – stock utterances –
 synthetic – unanalyzed chunks of speech – unanalyzed multiword chunks – units

Figure 4.1. Terms used to refer to various aspects of formulaic language (Wray, 2002: 9)

As the above figure clearly illustrates there is no shortage of names to refer to the notion of co-occurrence of lexical items. Weinert is of the opinion that “[w]hile labels vary, it seems that researchers have very much the same phenomenon in mind” (1995: 182). Wray (2002: 8), however, advocates caution in relation to the above claim and points out that researchers do not necessarily need to have exactly the same concept in mind. Such terms as those in Figure 3.1. are often used interchangeably, but since there is a danger that their respective scopes may vary from study to study, the need for explicit definition of the concept is an absolute necessity in any research project.

The term *formulaic language* or *formulaic sequence* is generally used to encompass a broader range of manifestations of the phenomenon than the term *collocation*. Formulaic language can be defined as follows:

...a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (Wray, 2002: 9)

The above definition, as Wray points out, is as inclusive as possible in order to accommodate various kinds of linguistic units that are considered formulaic.

The term *collocation* is more limited in terms of the length of the sequence to which it refers. It normally involves two linguistic units (e.g. Durrant, 2008; Gyllstad, 2009), with one of them appearing within a specified number of words to the left or right of the other. As outlined in Chapter 1, our project adopts the following definition of collocation:

two syntagmatically related and frequently co-occurring orthographic words, either adjacent or separated by a specified distance (...) where the meaning is evoked by the combination as a whole, sometimes requiring additional lexical elements for grammatical well-formedness and usage convention. (Gyllstad, 2009: 155)

The two linguistic units that constitute a collocation can be looked at from two perspectives, namely the part of speech they have or the frequency with which they occur. Within the former tradition, collocation studies focus on one type of collocation only, e.g. verb + noun phrases or adjective + noun phrases (Durrant, 2008; Gyllstad, 2005) or on a number of different types (e.g. Gitsaki, 1999). Within the latter tradition, collocation studies explore knowledge of high-frequency word combinations (e.g. Durrant, 2008) or they compare knowledge of high-frequency and low-frequency combinations. The above two traditions, i.e. the frequency-based approach to collocations and the phraseological (part-of-speech) approach, are discussed below.

4.3.2. Approaches to studying collocations and formulaic language

Two main approaches that have been used in research on collocation include the frequency-based approach of Sinclair and Hoey (2005) and the phraseological approach adopted by a number of contemporary writers such as e.g. Cowie (1981), Mel'cuk (1998) or Gitsaki (1999). The phraseological tradition is focused on the syntactic and

semantic analysis of collocations, and on compiling typologies of collocations on the basis of part-of-speech analysis of the constituent words. The frequency-based tradition, on the other hand, primarily takes account of collocation frequency and is less concerned with part-of-speech collocation types.

The frequency-based approach

Within the frequency-based approach collocations are generally regarded to be units made up of words that co-occur within a certain distance from one another with higher frequency than would be expected on the basis of coincidence or traditional linguistic rules (Durrant, 2008: 11). Within this tradition the work of Firth (1957 & 1974), Halliday (1961 & 1966) and Sinclair (1996, 1987a, 1987b, 1991 & 1996) is very often referred to.

According to the Firthian theory of meaning, as outlined in Section 4.2, meaning is context-dependent and can be established at “mutually congruent series of levels” such as collocation, syntax, phonology and phonetics (1974). Firth’s concept of *meaning through collocation* reflects the importance of arriving at word meaning through an analysis of its links with co-occurring words. This abstraction beyond words to sets of words is highlighted by the notion of semantic prosody and the idiom principle, also dealt with in Section 4.2. As for collocation types, Firth enumerates a number of different categories, such as “habitual” and “more restricted technical” for example, but he fails to provide more detailed descriptions of these categories. Within the syntactic level, Firth distinguishes between *collocation* and *colligation*, the latter referring to an association between a word (or a set of words) and a particular grammatical type. For example, “naked eye” tends to appear to the immediate right of the prepositions “with” or “to”, while “analysis” tends to occur to the immediate left of the preposition “of”. The term was later picked up by other researchers, including Halliday and Sinclair.

Building on Firth’s approach to the nature of lexical patterns in language, Halliday (1961) describes collocation as an association between linguistic units in a construction or sequence (syntagmatic axis) characterised in terms of the probability of co-occurrence of the units in question and the distance between the units. He argues that those patterns of language that cannot be accounted for by grammatical restrictions

“can only be accounted for in semantic terms” (1966: 149). For example, adjective such as “strong” and “high-powered” have roughly the same meaning, and yet they are known to enter into syntagmatic relationships with different nouns: “a high-powered car” and “strong tea”, but not the other way round (Halliday, 1966; see also Section 4.2.2 on near-synonymy). In the light of the above, the words “strong” and “high-powered” have the ability to ‘predict’ their own lexical environment in that they will co-occur alongside some nouns but not others (Halliday, 1966: 156). Three terms that Halliday uses to examine the nature of collocations are *node* (the lexical item under scrutiny), *collocate* (the co-occurring item) and *span* (the distance between the two items). These terms have since been used widely in operationalised definitions of L1 and L2 collocation research.

Sinclair’s work also originated from a Firthian view of collocation. He applied Firth’s ideas in practice to one of the largest projects in computational lexicography, the so-called COBUILD³⁸ project (described in e.g. Moon, 2009b and Sinclair, 1987b). The intention behind this project was to build a corpus of current English that would serve as a point of reference in the compilation of a monolingual English dictionary. In the 1970s relatively few linguists had access to corpus evidence, and the decision to build a dictionary on the basis of such evidence was considered innovative technologically at the time (Moon, 2009a: 4). During his work on the COBUILD project Sinclair expanded the Hallidayan notion of probability of co-occurrence by defining “the average ‘span’ within which any lexical item exerts a predictive influence” (Jones & Sinclair, 1974: 15). According to Sinclair’s calculations, a span of four orthographic words to the left and to the right of the node covers 95 per cent of the node’s collocates. Another major insight from the COBUILD project is that most frequent words in English “function as elements of structure” and collocate in a delexical rather than in a fully lexical sense (Renouf, 1987). In the phrase “have a bath” the verb “have”, for example, is used with an object to describe an action but it has very little meaning in itself. As Moon points out, “the only way to make distinctions in meaning or use within the major delexical verbs, such as *have*, *give*, and *take*, is to split according to the type of object collocate” (Moon, 1987: 94; original italics).

³⁸ The acronym COBUILD stands for Collins Birmingham University International Language Database.

As presented in Chapter 3, frequency counts extrapolated from a large general corpus can be used to explore knowledge and use of high- and low-frequency vocabulary in L1 and L2 contexts. Such frequency counts can also be used to establish patterns of co-occurrence of words in L1 and L2 texts. Within the frequency-based approach, frequency counts are thought to provide useful information about words that occur with the target item, and are treated as an indication of the level of strength of collocational links between words. Statistical procedures that are adopted to assess the strength of collocational links are outlined in Section 4.5.2.

The phraseological approach

The phraseological approach has largely been inspired by research carried out by Russian and European lexicographers (e.g. Cowie, 1998a; Cowie, 1998b), but it dates back further to the work of Palmer (1933 & 1934) in Tokyo in the early twentieth century. The work of Palmer and the Russian lexicographers developed alongside one another in an independent fashion. Russian phraseological theory that evolved between the 1940s and the 1960s was first presented to non-Russian speakers by Klappenbach (1968) and Weinreich (1969; cited in Cowie, 1998a: 4). Palmer's primary concern was categorising a list of collocations that could serve as a basis for the development of English language teaching materials and better learner dictionaries. Russian school lexicographers were also driven by the need to classify collocations in order to develop foreign language teaching materials and learner dictionaries. In his *Second Interim Report on English Collocations* Palmer (1933) enumerated a number of English collocation types that were based on a syntactic analysis of their constituent parts, and included *inter alia* verb + object collocations ("give trouble", "make a bed"), verb + preposition collocations ("argue with", "rely upon"), collocations marked by the absence of an article ("go to bed", "get hold of") and collocations without a break ("lighthouse", "birthday"). Russian school lexicographers developed other types of syntactic classifications of collocations, the most relevant of which are presented below.

Within the Russian school word units were divided according to *semantic opacity* (the degree to which words are used in their literal meaning) and level of *fixedness* (the degree to which elements of a phrase can be replaced by other words), but traditional Russian lexicology favoured the former criterion. Most of the early

schemes make a distinction between word-like units (e.g. “in the nick of time”, “break one’s journey”) and sentence-like units (e.g. “you don’t say”), the former functioning syntactically at or below sentence level and the latter functioning pragmatically as sayings and conversational formulae (Cowie, *ibid.*). Working at the level of word-like units, Vinogradov (cited in Cowie, 1998a: 4-5) distinguishes between phraseological fusions, phraseological unities, and phraseological combinations. *Phraseological fusions* (idioms) are semantically opaque units that are structurally fixed, e.g. “spill the beans”. *Phraseological unities* are partially opaque, as their originally transparent meanings have been extended metaphorically, e.g. “blow off steam”. *Phraseological combinations* (collocations) are word pairs that consist of two open-class words, one of which is used in its literal meaning and the other of which is context-dependent, e.g. the figurative meaning of the verb “meet” and the literal meaning of the noun “demand” in “meet the demand”. Amosova (also cited in Cowie, *ibid.*) subdivides phraseological combinations into phrasemes and phraseloids. *Phrasemes* are two-word phrases, in which the meaning of the non-literal word is found only when it is accompanied by a single collocater, e.g. “small (trivial) talk”, “small (of low value) change” and “small (early) hours”. In *phraseloids* the meaning of the non-literal word is found only when the word is accompanied by a number of collocates, e.g. “pay one’s respects” and “pay a compliment”.

Modern phraseologists have been heavily influenced by the above-mentioned classifications of collocations. Cowie (1994), for example, divides phrases into *formulae* (corresponding to sentence-like units presented above), *pure idioms* (phraseological fusions), *figurative idioms* (phraseological unities), and *restricted collocations* (phraseological combinations). Howarth (1998: 35) specifies six levels of fixedness of collocations in relation to verb + noun phrases. Level 1 refers to so-called free collocations, Levels 2, 3 and 4 refer to restricted collocations, while Levels 5 and 6 refer to idioms. Level 1 is characterised by a restriction on the choice of the verb but relative freedom of replacement of the co-occurring noun, e.g. a more restricted verb set (“compare”, “emphasise” and “influence”) plus a relatively non-restricted noun set (“behaviour”, “groups”, “levels”, “culture”, “content”, etc). Level 2 is characterised by some degree of substitution of both elements, e.g. a small verb set (“introduce” and “bring forward”) plus a small noun set (“a bill” and “an amendment”). At Level 3 a small number of synonymous verbs are used in a specific sense with one noun only

(e.g. “take heed” or “pay heed”). At Level 4, one verb is used with a small range of nouns (e.g. “give credit to somebody” or “give preference to something”). At Levels 5 and 6 there is a complete restriction on the substitution of both elements (e.g. “draw a line”). As the above classification illustrates, Howarth considers the degree of fixedness to be the primary criterion for identifying and classifying collocations, unlike Vinogradov (cited in Cowie, 1998a: 4-5) for whom semantic opacity was the primary focus.

Another typology of collocations within the phraseological tradition is the framework utilized in *The BBI combinatory dictionary of English* (Benson et al., 1986). This typology divides collocations into grammatical and lexical³⁹, the former consisting of an open and a closed class word (e.g. adjective + preposition) and the latter composed of open class words (e.g. verb + noun). Eight main types of grammatical collocations (G1-G8, e.g. preposition + noun) and seven types of lexical collocations are identified (L1-L7, e.g. adjective + noun; see Table 4.1 below). Some of the categories represent only one type of construction (G1, i.e. noun + preposition) while others include a number of subtypes (G8, i.e. 19 verb patters).

Table 4.1. Selected collocation types in the BBI typology (Benson et al., 1986)

Grammatical types of collocation (G1-G8)	Lexical types of collocation (L1-L7)
G1 - noun + preposition : <i>apathy towards</i>	L1 - verb (meaning creation, activation) + noun/pronoun/prepositional phrase : <i>come to an agreement</i>
G2 - noun + to-infinitive : <i>to have the right to do it</i>	L2 - verb (meaning eradication, nullification) + noun : <i>reject an appeal</i>
G3 - noun + that-clause : <i>an agreement that she would do it</i>	L3 - adjective + noun : <i>strong tea</i>
G4 - preposition + noun : <i>by accident, in advance</i>	L4 - noun + verb : <i>blood circulates</i>

The tests of English collocations developed by Gitsaki (1999) and used in the current study comprise different types of grammatical and lexical collocations from the BBI typology (see Appendix E, Tables A.1-A.3).

4.3.3. Selected functions of collocations and formulaic language

Collocations are believed to have a number of functions. They are used as a fluency device when processing pressures are high. They allow the speaker to manipulate the environment and express his or her perceptions, attitudes or emotions

³⁹ The distinction between lexical and grammatical collocations parallels Firth’s distinction between collocation and colligation.

(Wray, 2002). They also help the speaker to achieve his or her intended communicative goal in specific situations, and allow the speaker to express the content that he or she would otherwise be unable to construct from his or her interlanguage rules (Nattinger & DeCarrico, 1992). The main kinds of functions collocations perform can differ from one discourse to another; e.g. in social interaction collocations are believed to perform mainly an interactional role, while in a written context their transactional/ referential function becomes more apparent (Moon, 1992). They can also be used creatively in writing to create a specific stylistic effect (Partington, 1998).

The most relevant function of collocations and formulaic language from the perspective of the current research project is their role in alleviating the speaker's and the hearer's processing effort. It has been shown that when processing pressures are abnormally high, very little non-formulaic language is produced (Wray, 2002: 17). As illustrated by Foster (2001), when native speakers are asked to participate in interactive tasks, such as for example a discussion, without any time allowed for planning, they tend to rely more heavily on formulaic language. An example of this is given by Nattinger and DeCarrico (1992). During a recorded meeting of the Portland State University Faculty Senate in 1989 when a question of faculty salaries cropped up, a representative from the Governor's office had this to say:

Now having said that, I want to share with you what is on my own mind and, at the time, I want to make it perfectly clear that, in this context at any rate, I'm one of those who thinks that – and this is purely a personal opinion, and one that in fact I'm not sure that the Governor would share – to make it clear that in this climate there may be a general need for reduction. (PSU Faculty Senate, quoted in Nattinger & DeCarrico, 1992: 74)

As the above quote shows, apart from giving speakers more time to plan the discourse, formulaic language also allows them to manoeuvre themselves out of a potentially difficult and unpleasant situation. The fixed chunks with which the above extract is packed are meant to allow the speaker to distance himself or herself from the hot issue of salary cuts, and appear less responsible for the bad news being passed on.

In everyday communication L2 formulaic language and collocations allow the speaker to express ideas of more lexical complexity than would otherwise be possible. L2 learners sometimes pick up longer or shorter prefabricated chunks that remain

unsegmented and are not recognized as “units in linguistic analysis” (Bolander, 1989: 81). Even though the students’ awareness of the internal structure of these units might be very limited or non-existent, the use of these phrases allow them direct entry into communication to express what they would otherwise be unable to construct creatively from their interlanguage rules (Nattinger & DeCarrico, 1992: 27-28). “I’m finished”, for example, can be used by a non-native speaker – without much awareness of what each constituent lexical item means – to indicate e.g. the wish to leave a place (Weinert, 1995: 186). An expression of that sort can at a later point be broken down into smaller patterns, though. Brown presents an account drawn from a student learning Japanese of how unsegmented items can undergo the process of segmentation:

Hearing again and again the question *Kore wa nan desu ka?* (What is this?) but never seeing it printed I conceived of *korewa* as a single word; it is spoken without pause. Some lessons later I learnt that *wa* is a particle (...). Interestingly enough I did not, at once, reanalyze my word *korewa* and such others as *sorewa* and *arewa* into noun and particle forms. I did not do that until I started to hear such object forms as *kore o* and *sore o* and *are o* in which *o* marks the direct object. Then the truth dawned on me... (Brown, 1973: 5; original italics)

Most L2 learners could undoubtedly identify with the above anecdote. However, it should also be pointed out that there are others too who, despite their large quantity of accumulated L2 formulas, do not seem to be that efficient in the segmentation process and exhibit signs of having a fossilized grammatical system (Schmidt, 1983). The issue whether the use of formulaic language can contribute to the development of rule formation in child L1 acquisition and adult L2 learning is addressed in the coming section.

4.4. Collocations and formulas in the theory of language learning

From a theoretical perspective the notion of collocation has contributed to shaping various language models – those of language acquisition (Ellis, 2001; Tomasello, 2003; Wray, 2002), fluency and nativelike selection (Pawley & Syder, 1983) and language processing (Ellis, 2002; Jiang & Nekrasova, 2007; Schmitt, 2004). In this section two influential models of language acquisition are considered in turn, that of Ellis (1996) for L1 acquisition and that of Wray (2002) for the L2 context. The

notion of chunking in sequence learning is also brought to the readers' attention, as it is posited to be an important psychological mechanism that helps to explain the phenomenon of collocation (Ellis, 1996 & 2001).

The subsequent section presents Durrant's (2008) empirical exploration of L2 collocation learning in adults that extends the applicability of the Ellis model from the L1 to the L2 context.

4.4.1. L1 collocation learning

The Ellis (1996) model of L1 collocation learning

The Ellis (1996 & 2001) model of L1 language learning is driven by the principle that sequence learning lies at the very heart of the process of language learning: "language learning involves learning sequences of words (frequent collocations, phrases, and idioms) as much as it does sequences within words" (Ellis, 2001: 45-46).

Sequence learning is closely tied to the notion of chunking. The term *chunking* was first introduced by George Miller (1956) in his review of STM, in which he argued that the span of STM remains more or less constant at the level of approximately 7 items, regardless of how much information is encoded as one chunk. The implication was that by coding simple items into more complex chunks we could substantially increase the amount of information our memory can hold. The Ellis model of language learning, based on the very same principle of chunking, posits that smaller chunks are available for re-coding into bigger chunks, and bigger chunks can in turn be combined into yet larger chunks. According to this account, learning vocabulary involves sequencing at a sub-lexical level, that is at the level of syllable structures and phonotactic sequences. Learning discourse involves the sequencing of phrases and collocations, while learning grammar involves abstracting structural similarities from "previously experienced utterances which share structural and functional similarity" (ibid.: 49).

The acquisition of L1 word forms, collocations and grammatical aspects occurs, Ellis (2002) maintains, mainly through implicit processes of analysis of sequence information. Acquiring a language in that way provides the language user with a vast

amount of predominantly unconscious knowledge about sequence probabilities in his/her language (Section 4.2). However, the acquisition of collocations, Ellis posits, involves more than creating purely formal associations, meaning being an essential part of the process:

The naturalistic learner doesn't care about (...) theories of grammar or whether words or morphemes are the atomic units of language. From a functional perspective, the role of language is to communicate meanings, and the learner wants to acquire the label-meaning relations. (Ellis, 2001: 40)

The significance of meaning in sequence learning is reflected in the fact that when a chunk activates some meaning representation it becomes more salient to the language learner and more likely to be learnt compared to other sequences of sounds that do not activate such representations. The mapping of form to its communicative function, Ellis (2005) maintains, requires the establishment of cross-modal links, and that in turn requires a conscious focus of attention. Once the form-meaning association is explicitly made, the chunk is subject to implicit processes and frequency effects.

Formulas in child L1 learning

Proponents of formula-based approaches maintain that the learning of fixed formulas is an essential aspect of the learning process, and that it leads to the development of creative aspects of language (e.g. Nattinger & DeCarrico, 1992: 114-116; Tomasello, 2003: 305-307). This section outlines a number of key points related to formula-based learning in L1 acquisition first (e.g. Lieven & Tomasello, 2008; Peters, 1983; Tomasello, 2003), and then addresses the general patterns that can be observed in L2 learners.

One of the key assumptions of usage-based L1 acquisition models is the link between the route of language acquisition and the nature of input the child receives. According to these approaches, children's acquisition of particular lexical sequences largely depends on the kind of language a given child is exposed to. The input children receive is like raw material which they use as a basis for constructing their linguistic environment. It is fairly obvious, though, that different children are exposed to different types of constructions. In the initial stages their language is likely to reflect simple constructions they encounter in speech directed at them (Cameron-Faulkner, Lieven &

Tomasello, 2003; De Villiers, 1985; Hoff & Naigles, 2002). Eventually, however, they end up “with fairly similar (though not identical) linguistic inventories” (Tomasello, 2003: 110).

The proposition of usage-based models that the development of the language system is hugely dependent on the nature of the received input has been claimed to stand in contrast with Chomskyan concept of an innate language acquisition device. In Barlow and Kemmer’s (1999) formulation:

...if instances of use are the prime input driving the system’s formation, then positing genetically-specified guiding linguistic structures is unnecessary.
(Barlow & Kemmer, 1999: xii)

The above quotation points to an important aspect of usage-based accounts, namely the assumption that the process of forming language happens not via innate principles and parameters but through an interaction between the environment and the mind of the language user (MacWhinney, 1999). Linguistic representations are seen as cognitive routines rather than fixed entities stored in any particular neural connection. They are described as “nothing more than recurrent patterns of mental (ultimately neural) activation” (Barlow & Kemmer, 1999: xii), such that when there is no processing, the information that linguistic units represent resides in connectivity patterns from previous activations. In contrast with the idea of an innate acquisition device, which operates on different principles from the rest of the mental system, the process of language acquisition and non-linguistic types of learning are seen on this view as being governed by the same psychological mechanisms.

Many researchers have noted that there is great variation in the extent to which children focus on learning lexical chunks as opposed to learning individual words. Peters (1977) identifies two general learning tendencies in this connection, namely *gestalt* and *analytic* learning strategies. The former involves proceeding from the whole to the parts, while the latter involves proceeding from the parts to the whole. Having stated that the *gestalt* strategy is probably quite widespread and that it needs to be taken into account in theories of language acquisition, she also admits that:

...there is probably a continuum of children, varying from those who are very Analytic right from the beginning, through those who use mixes of Analytic

and Gestalt speech in varying proportions, to those who may start out with a completely Gestalt approach and have to convert slowly and painfully to a more Analytic approach to language. (Peters, 1977: 570-571)

What Peters labels as *analytic* and *gestalt*, other researchers refer to using other terms, such as e.g. *referential* and *expressive* (Nelson, 1981), or *noun-lovers* and *noun-leavers* (Horgan, 1980). The concept behind these labels is generally believed to be the same, though, namely that there are two extreme points of “continua along which most children are ranged, very few being clearly at one pole or the other” (Peters, 1983: 20).

A number of factors are believed to be responsible for the size of lexical units that children extract from the input (Peters, 1983: 20-34), ranging from the child’s communicative needs to birth order. As for the child’s communicative needs, the need to label various things around them can encourage children to extract words to use as labels from adult speech, e.g. *kittycat*, *rocking horse*, etc. On the other hand, the need to communicate in a social context can drive them to extract multi-word formulas necessary to conduct those interactions, e.g. *give me that*, *you do it* or *nice to see you* (ibid.: 22-23). The size of units children extract may also be influenced by other factors such as culture, social class and birth order (Montgomery, 1977; Schieffelin & Eisenberg, 1981). Differences in the kind of speech that adults from various social and cultural backgrounds direct at children may result in different extraction strategies. Similarly, the kind of input first-born children receive as compared to that directed at their younger siblings may play a role as well. The third of the factors – “a culturally determined set of expectations regarding appropriate language use” (Peters, 1983: 28) – may also have an impact on the type and size of the earliest extractions and on children’s perceptions of appropriate occasions of use. Two other factors that may affect early extraction strategies are learning disabilities and differences in STM capacity (see Section 4.6).

4.4.2. L2 collocation learning

The question arises as to whether formulas perform a similar role in adult L2 learning to that played in child L1 learning. The two aspects of formula learning addressed in relation to adults include the ability to extract lexical information from

learnt phrases (Bolander, 1989; Myles, Hooper & Mitchell, 1998) and the ability to learn collocations from L2 input (Durrant, 2008).

The Wray (2002) model of L2 collocation learning

The Wray (2002) model addresses the issue of L2 collocation learning in adults. Having undertaken an extensive overview of studies that investigate the acquisition of formulaic language in adult L2 learners, Wray concludes that in the early stages of language learning formulaic sequences are relatively easy to acquire but “by the time the learner has achieved a reasonable command of the L2 lexicon and grammar, the formulaic sequences appear to be lagging behind” (2002: 182). In her book *Formulaic Language and the Lexicon* Wray (2002) proposes an L2 acquisition model according to which adult L2 learners tend to retain individual words but do not normally acquire collocations from the language input they receive (2002: 206-209). If her model is valid, adult L2 learners, unlike native speakers, break up the input they encounter into separate lexical items and fail to retain the information about how these words are combined:

...the characteristic way for formulaic sequences to be handled by post-childhood language learners is that they are broken down in order to access the lexical constituents, which are then stored separately. The information about (...) how they were linked is not retained. (2002: 210).

The above quotation would suggest that there is a fundamental difference between the way children learn collocations from L1 input and the way adults learn them from L2 input. According to Wray, there are a number of social and cognitive factors that account for this difference. Some of the reasons she enumerates in this connection include teaching methods, learning environment, peer pressure and communicative needs. She also draws attention to adults’ increased awareness of “the word as a possible unit of linguistic processing (a natural product of (...) being literate)” and how that might result in adults being more predisposed to acquire a lexicon of word-sized units (ibid.: 205-206). One cannot reject out of hand the possible validity of this latter argument. However, the idea that all L2 learners fail to pay attention to information on L2 word combinations is counter-intuitive (Williams & Lovatt, 2005: 185). One would expect there to be some variability among that group of

learners in the extent to which they commit multi-word units to memory (Dörnyei, Durow & Zahran, 2004). One could also expect that individual differences in the ability to commit such units to memory could be related to variability in the efficiency of chunking these phonological units in STM. Schmitt et al. (2004b), for example, suggest that limited memory capacity may be one of the factors that explain why some of their subjects “did not seem to have the recurrent cluster available as formulaic sequences, and so tried to generate a sensible reconstruction based on (...) key words” (ibid.: 140). The issue of memory capacity in the learning of collocations is discussed further in Section 4.6 later in this chapter.

Formulas in adult L2 learning

According to Nattinger & DeCarrico (1992), in relatively naturalistic contexts adults, just like children, go through a stage in which they put together memorized chunks of lexis to produce socially acceptable responses and then later “on analogy with many similar phrases, they break these chunks down into sentence frames that contain slots for various fillers” (ibid.: 114-115). Similarly, Lewis (2002: 95-98) posits that lexical phrases are broken down by adult learners into their constituent parts in order to extract underlying rules in a similar manner to the one described in relation to children (see also Ellis, 2003: 72). Collocations could, therefore, be viewed as a means of gaining a foothold in the grammar and the lexicon and as a means of facilitating the formation of a creative and nativelike L2 system (e.g. Ellis, 2003: 73-74; Nattinger & DeCarrico, 1992: 11-17). Wray (2000: 472), however, advocates caution with respect to the claim that adult L2 learning reflects the same formula-based processes as child L1 learning does, and states that “there is very little evidence that adult learners naturally extrapolate grammatical or lexical information from larger strings”.

There are a number of factors that may result in different patterns of collocation learning in L1 and L2. In addition to those presented earlier in the context of Wray’s (2002) model of L2 learning, these factors include conceptual development, type of language input and transfer from L1 (Ellis, 2003). With regard to conceptual development, child L1 learning is accompanied by the development of knowledge of the world, while adult L2 learning takes place in relation to pre-existing conceptual knowledge. Adults possess more sophisticated analytic processes that enable them to “treat language as an object of explicit learning” (ibid.: 72) and apply conscious

problem-solving strategies as they see fit. Apart from having a mature cognitive system at their disposal, adults are also equipped with a fully-developed L1 system. Children build lexically specific patterns and then develop the knowledge of abstract syntactic categories, while adult L2 learners already possess the knowledge of abstract categories and of the way they are combined in their L1. This, as Ellis (2003: 72) claims, “may guide creative combination in their L2 interlanguage to variously good or bad effects.” What might increase the operation of the cross-linguistic factor is a pressing need to communicate a message that is beyond the L2 capacity of the adult learner. Being unfamiliar with the relevant formulas, the learner might also resort to non-linguistic means or refrain from communicating the intended message at all (Wray, 2002: 175).

A recent laboratory-based study by Durrant (2008: Chapter 5) shows that adult L2 learners of English have the capacity to retain some information about word combinations from L2 input. The non-native proficient speakers of English who participated in the experiment were presented with sentences containing English adjective + noun combinations⁴⁰ in a number of experimental conditions. In the testing session an adjective from one of the adjective + noun combinations was first presented, and then the stem of the corresponding noun was displayed on the screen (e.g. ‘FL_ _’ for *flat*). The subjects were not explicitly told about the connection between displayed adjectives and nouns. The level of recall for the nouns that were part of the target adjective + noun combinations was significantly higher than the level of recall for other nouns. In the light of the above, Durrant argues that the fact that adult learners in his study could “automatically retain a memory of collocational chunks from the language to which they [were] exposed” speaks in favour of applying Ellis’s model to adult L2 collocational learning, rather than Wray’s (*ibid.*: 165).

Although Durrant’s data suggest that adults may indeed have the skill to pick up collocational chunks from the input they receive in different experimental conditions, Wray (2002: 182) may also have a point in saying that, even when L2 learners have sufficient vocabulary to be proficient in L2, their collocational knowledge lags behind. As mentioned in Chapter 2, the strength of links between words comprising a collocation could be weaker as compared with the strength of links between syllables

⁴⁰ To avoid possible confounding effects of pre-existing collocational links between selected nouns and adjectives, Durrant ensured that the target noun + adjective combinations appeared with zero or very low frequency in the BNC

comprising a word – due to the fact that the amount of exposure to a word is higher than the amount of exposure to a collocation in which that word features. As a result, it may take longer to strengthen links between words comprising a collocation than it does to strengthen the links between syllables comprising a word.

4.5. Measurement of collocation knowledge

L2 collocation knowledge can be tapped in a number of different ways, just as L2 vocabulary knowledge can (Chapter 3, Section 3.4). Methods used to elicit information about L2 collocation knowledge include receptive and productive collocation tests (Bahns & Eldaw, 1993; Biskup, 1992; Bonk, 2001; Eyckmans, 2009; Gitsaki, 1999; Gyllstad, 2005; Gyllstad, 2009; Revier, 2009), recordings of non-native speech or writing (e.g. Durrant & Schmitt, 2009; Gitsaki, 1999; Granger, 1998; Nesselhauf, 2005), and laboratory-based methods, such as eye-tracking or priming (e.g. Durrant, 2008; Ellis & Sinclair, 1996).

This section presents a number of key receptive and productive tests of collocation knowledge, and evaluates some statistical methods used to establish the strength of collocational links between words. The tests of collocation knowledge described in below include Gitsaki (1999) Collocation Tests, utilized in the current study, and other types of collocation tests available at the time when our project was in its design stage, e.g. the COLLEX Test (Gyllstad, 2005 & 2009). This section also contains a description of selected statistical measures that calculate the strength of collocations (i.e. raw frequencies, Mutual Information (MI) score, Mutual Information cubed (MI3) score or t-score).

4.5.1. Receptive and productive tests of collocation knowledge

A considerable number of different tests have been employed in L2 collocation knowledge research. Some of these collocation tests have been extensively validated with a view to being used across a range of studies⁴¹ (e.g. Bonk, 2001; Gyllstad, 2005 & 2009). Other collocation tests have been created for the purpose of one particular study, and have not been re-used in other research projects (e.g. Bahns & Eldaw, 1993;

⁴¹ None of these tests, however, has been employed as widely in collocation studies as e.g. the Vocabulary Levels Test in vocabulary breadth studies (Chapter 3).

Biskup, 1992; Farghal & Obiedat, 1995). More often than not, researchers choose to create their own tasks to tap the specific aspect of collocation knowledge they are interested in exploring (e.g. Leśniewska & Witalisz, 2007). This section critically reviews a number of the test formats used in collocation research, and presents key collocation tests that make use of the formats in question. It also presents the formats employed in our research project, i.e. a forced-choice preference format, acceptability judgement tasks (the Progress Collocation Test, Chapter 6), and a fill-in-the blank procedure (Gitsaki Collocation Test and the Progress Collocation Test).

The most frequently used formats of collocation tests include multiple-choice questions (e.g. Schmitt, 1999), forced-choice preference tasks (e.g. the COLLEX format in Gyllstad, 2005), acceptability judgements (e.g. Leśniewska & Witalisz, 2007), fill-in-the-blank procedures (e.g. Gitsaki, 1999) or translation (e.g. Bahns & Eldaw, 1993; Biskup, 1992). An examination of the formats used in collocation knowledge research (Ellis & Barkhuizen, 2005; Rivers & Temperley, 1978) point to the fact that the test formats used in collocation research have their strengths and weaknesses, just as the formats employed in vocabulary breadth testing do. We shall start by presenting two methods that are particularly popular in measuring receptive knowledge of collocations, namely multiple choice-questions and acceptability judgement tasks.

Multiple-choice questions involve choosing one answer from a number of options provided, while acceptability judgement tasks involve judging whether the presented phrases/sentences are correct or not. Multiple-choice questions normally provide the test-takers with more than three options to choose from. When only two options are included, as in our Vocabulary Progress Test (Chapter 6), the format is referred to as a forced-choice preference test (see the COLLEX test in Gyllstad, 2005 below). Even though multiple-choice questions and acceptability judgement tasks sometimes come with the additional “not sure” option, some learners may resort to guessing or other compensatory strategies to produce an answer instead of admitting that they do not know the answer. In order to limit guessing opportunities in acceptability judgement tasks the test-takers can additionally be asked to carry out further operations on the test phrases/sentences, that is to indicate which part is incorrect or to write out the phrases correctly, as was done in the course of our research project. Both tasks can also be administered in a speeded (i.e. pressurised) format, in

which learners are given a specific (limited) amount of time to complete the task. One has to bear in mind that although efforts can be made to limit guessing opportunities, guessing itself cannot be eliminated completely (see Chapter 3, Section 3.4. for more on guessing in vocabulary breadth testing).

An example of the use of a forced-choice preference collocation task is provided by a test called COLLEX, designed and validated by Gyllstad (2005). The test measures receptive knowledge of 50 decontextualized verb + object combinations⁴². In justifying his preference for employing decontextualized items, Gyllstad refers to Cameron's (2002) statement about the processing of decontextualized language, namely that:

...we can surmise that decontextualized presentation of a word in a test does not imply that a testee makes sense of the test word in a 'decontextualized' mental void. Rather, the recognition process may activate recall of previous encounters and their contexts (Cameron, 2002: 151)

Gyllstad also refers to the fact that the use of decontextualised items in collocation testing limits the amount of reading and inferencing that would otherwise be taking place. In this test format, test-takers are presented with juxtapositions of a real collocation (e.g. "make a mistake") and a pseudo-collocation (e.g. "do a mistake"), and are instructed to select the one they think is real. The test-takers are also asked to indicate whenever they resort to guessing by ticking a specially-designated box. In the process of the test construction the BNC was consulted to rule out the possibility that some of the pseudo-collocations were actual lower frequency collocations. A relatively high overall reliability measure ($\alpha = .91$) was obtained for a revised version of the COLLEX Test. However, since risk-taking tendencies and resultant guessing may be problematic in this type of test format, it would seem appropriate to combine the forced-choice preference format with other testing techniques in order to obtain a more reliable measure of learners' receptive knowledge of collocations. The COLLEX Test focuses on one type of lexical collocation only, and since our research project aimed at measuring a number of types of lexical and grammatical collocations, the COLLEX Test was deemed inappropriate for our research purposes.

⁴² Other noun + object tests that have been recently developed include e.g. the Discriminating Collocations Test (Eyckmans, Boers & Stengers, 2006; Eyckmans, 2009) or the CONTRIX Test (Revier, 2009).

Collocation tests that employ a fill-in-the-blank procedure involve eliciting predetermined linguistic features in learners' production (Ellis & Barkhuizen, 2005: 30). Since they serve as a tool to elicit specific information from the students, it has to be remembered that they only document the kind of constructions learners can produce under the conditions of the testing procedures. The patterns of results obtained in that manner may or may not represent the kind of language the students would produce under more natural conditions of use (see for example the study carried out by Burmeister & Ufert, 1980).

Two sections of Bonk's (2001) Collocation Test make use of a fill-in-the-blank format. Bonk's Collocation Test includes the following task formats: 1) 17 fill-in-the-blank sentences probing knowledge of selected verb + object collocations, 2) 17 fill-in-the-blank sentences measuring knowledge of selected verb + preposition combinations, and 3) 16 multiple-choice items testing familiarity with figurative-use-of-verb phrases. Some examples from subtests 1, 2 and 3 are presented below:

Subtest 1: Punk rockers dye their hair red and green because they want other people to _____ attention to them.

Subtest 2: Many of the birds in the area were killed _____ by local hunters (to exterminate).

Subtest 3: a. Are the Johnsons throwing another party?
b. She threw him the advertising concept to see if he liked it.
c. The team from New Jersey was accused of throwing the game.
d. The new information from the Singapore office threw the meeting into confusion.
(more than one correct answer possible – our comment)

The test was piloted on 10 native speakers, and then redrafted and validated on 10 native speakers other than those recruited for the first validation. Subsequently, it was validated on 98 mostly East-Asian adult learners of English from the University of Hawai'i, whose level of proficiency varied from intermediate to advanced. The reported Kuder-Richardson 20 analysis of internal consistency for the whole test yielded a high reliability value of .83, but the second subtest, involving preposition + verb constructions, was found to have an unacceptably low reliability coefficient of .47. Having conducted statistical analyses such as item facility and item discrimination analysis, Item Response Theory analysis and generalizability analysis, Bonk concludes that his test is a reliable measure of collocational proficiency, apart from subtest 2, which, as Bonk suggests, could be discarded in favour of extending the other two

subtests. There are a few minor drawbacks to the study that Bonk himself pointed out in his paper (e.g. cultural bias in some of the collocation test prompts – p. 138), but nonetheless, it needs to be highlighted that Bonk's rigorous approach to test design and item validation sets a very good example for SLA researchers working in the field of collocations. The fact that Bonk's Collocation Test taps knowledge of more than one type of collocation made it more suitable in this respect for the current study than the COLLEX Test. However, Bonk's Collocation Test was developed with intermediate and advanced learners in mind, and was not, therefore, deemed appropriate for our subjects.

Another set of tests that employs the fill-in-the-blank format are Gitsaki (1999) Collocation Tests. As discussed in detail in Chapter 6 (Section 6.3.4), Gitsaki developed three collocation tests on the basis of classroom materials used in one of Greece's junior high schools. Each of the tests was designed for a different English proficiency level: elementary, intermediate and advanced. Gitsaki Collocation Test for beginners consists of 50 test items. 39 of these items can be classified according to the BBI typology (Benson et al., 1986) as four types of grammatical collocations (24 test items) and five types of lexical collocations (15 test items). The remaining 11 test items are phrasal verbs and thus are not listed in the BBI typology. Gitsaki Collocation Test for intermediate students consists of 65 test items. 52 items can be classified as four types of grammatical collocations (36 test items) and four types of lexical collocations (16 test items) according to the BBI typology. The remaining 13 test items are phrasal verbs.

The fact that Gitsaki Collocation Test taps knowledge of a number of types of lexical and grammatical collocations made it a good candidate for inclusion in our instrument battery (see Chapter 6, Section 6.3.4 for a more detailed description of the test). Since the current study did not aim to make direct comparisons between A2 (elementary) and B1 (pre-intermediate) students but focused on within-group differences, we felt that using two collocation tests specifically developed for each level instead of one test was an acceptable option. However, Gitsaki Collocation Tests are not free from drawbacks. Unlike Bonk (2001) or Gyllstad (2005), Gitsaki (1999) unfortunately did not report reliability estimates for the tests and item analysis results, even though she had collected data from a substantial number of students (91

beginners, 94 intermediate students and 90 upper-intermediate). Also, some fill-in-the-blank sentences on the test were designed in such a way that they allowed the insertion of more than one collocate of the target word. Gitsaki's account does not make clear whether she accepted alternative answers as correct or rejected them as incorrect in the course of her study. In the light of the above, it was necessary for us to pilot Gitsaki Collocation Tests (for beginners and intermediate students) on A2 and B1 Polish learners of English before deciding whether they would be suitable for the subjects in our main study or whether a general collocation test would need to be developed from scratch (see Chapter 5, Section 5.3.2).

4.5.2. Statistical measurements of collocational links between words

Raw frequencies of collocations

In corpus linguistics the term *frequency* denotes “the arithmetic count of the number of linguistic elements (i.e. tokens) within a corpus that belong to each classification (i.e. type) within a particular classification scheme” (McEnery et al., 2006: 52). Raw frequency is the most direct quantitative datum that can be extracted from a corpus. In studies like ours where only one general corpus (the BNC) is accessed to obtain the frequencies of the collocations under scrutiny, raw frequencies can be used to compare the number of occurrences of different lexical patterns within the same corpus. In studies in which two corpora or sections of the same corpus of a markedly different size are compared, raw frequencies are generally normalised to a common base.

The raw frequencies of collocations can be extracted from a language corpus with the use of query language, such as that provided by Hoffmann et al. (2008) for the BNCweb, a user-friendly web-based tool that supports analyses of the BNC (see Chapter 3, Section 3.3.2 for more information on the BNC). Raw frequencies can also be calculated in the BNCweb and other applications of the type by specifying the node word and the span within which instances of all words co-occurring with the node are to be analysed. The second type of count ignores sentences boundaries, and hence some of the co-occurring words may occur within the specified span by chance. Our own

study utilised the former method, i.e. query language, to avoid problems with sentence boundaries.

Within the frequency-based approach collocations are typically divided on the basis of their frequency counts into a number of frequency bands, e.g. low-, medium- and high-frequency (e.g. Siyanova & Schmitt, 2008). Raw frequencies can be used to obtain information about the relative frequencies of occurrence of word combinations in a given corpus (e.g. Durrant, 2008; Gyllstad, 2005; Schmitt et al., 2004a), but they cannot be used to make inferences about the strength of the word combinations. The reason why raw frequencies should not be used as a criterion for establishing the *strength* of collocational links between words is that it could lead to pairs of high frequency words that have weak collocational links being mistakenly classified as strong collocations. For example, according to the BNC counts, “sweet shop” occurs more often than “sweet peas”, which could be taken as an indication that the former has stronger collocational links than the latter. However, this is not the case when the total frequencies of the words “sweet”, “shop” and “peas” are brought into the picture. The noun “peas” occurs 612 times in the BNC and in 49 cases with the adjective “sweet”. The noun “shop” occurs as many as 10,066 times in the BNC, but only 50 times with the adjective “sweet” (McEnery et al., 2006: 215). Considering total frequencies of the nouns “peas” and “shop”, the former but not the latter has a strong preference to collocate with the adjective “sweet”. The use of raw frequencies should, therefore, be accompanied by other methods of identifying collocations⁴³.

Measures of collocational strength and certainty

Several methods have been proposed as a means to identify word pairs that are collocations in the frequency-based sense. The MI score is a statistical formula borrowed from information theory to measure collocational strength (McEnery et al., 2006: 56). It is calculated by dividing the observed frequency of the co-occurring word within the specified span by the expected frequency of the co-occurring word in the corpus as a whole. The result is then converted to base-2 logarithm (Hunston, 2002: 70). The higher the MI score, the stronger the link between the two lexical items under scrutiny is believed to be. The closer to zero the MI score is, the more likelihood there

⁴³ Our own study, apart from using raw frequencies, also involved consulting collocation dictionaries and calculating MI3 scores (see Chapter 1 for detailed operational definitions of the construct of collocation).

is that the items co-occur by chance. An MI score of 3 or more is taken to be significant (Hunston, 2002: 71). Since the MI score is based on the independent relative frequency of two lexical items in a corpus, some connections between items may be estimated to be strong due to the nature of the corpus in question.

To illustrate the extent to which independent relative frequencies can influence the MI score, Hunston (2002: 71-72) presents an analysis of the noun “gaze” and words co-occurring with it in the Bank of English (a general corpus). The adjective “baleful” has an MI score of 8.7, which places it among 15 top collocations of the node “gaze”. Due to the fact that “baleful” is a very infrequent word and occurs only six times in the corpus, even such a low co-occurrence of the two words is rendered as significant. The adjective “baleful” is generally linked with the noun “gaze”, but perhaps not with the strength indicated by its MI score. Other, rather unexpected items feature on the list of 15 top collocations of the noun “gaze”, e.g. “Ulysses” (MI of 9.1) and “Angelopoulos” (MI of 12.1). The presence of these two items on the list can be accounted for by the fact that the Bank of English contains reviews of the film “Ulysses Gaze” directed by Angelopoulos. As the case of “Angelopoulos” and “gaze” illustrates, “knowing the strength of the collocation is not always a reliable indication of meaningful association” (Hunston, 2002: 72). To give more prominence to high frequency words, the calculation of a t-score is very often used.

T-score is a measure of certainty of collocation, and it reflects the amount of evidence there is in a given database for the co-occurrence of two words. T-score is calculated by subtracting the expected from the observed frequency of the co-occurring word⁴⁴, and dividing the result by the standard deviation. A t-score of 2 and higher is normally considered as significant (Hunston, 2002: 72). To use “his gaze” as an illustration of the difference between t-score and MI score, in the Bank of English the phrase is at the very top of the list of collocates of the noun “gaze” with a very high t-score of 25.2. This indicates that in this general corpus “his” and “gaze” co-occur with very high frequency. The phrase “his gaze” does not have a very high MI score, though, because “his” co-occurs with many other words apart from “gaze”. In other words, there is a huge amount of evidence for co-occurrence of the two items (reflected

⁴⁴ Just as in the case of MI score, “the observed” is the actual frequency of the co-occurring word within the specified span, while “the expected” is the expected frequency of the co-occurring word in the corpus as a whole.

in their high t-score), but the collocational link between them is rather weak (reflected in their low MI-score). The top 15 collocates of the noun “gaze” from the point of view of t-score provide information mainly about its grammatical behaviour, and include many prepositions and possessive forms, while the top 15 collocates from the point of view of MI score provide more information about its lexical behaviour and more fixed occurrences.

As illustrated above, MI scores and t-scores tend to emphasise different sets of word pairs. MI scores give more prominence to less frequent word pairs, whose constituent words are not often found apart. T-scores tend to favour very frequent collocations, and in that respect are very similar to rankings based on raw frequency. For that reason “pairs like *good example*, *long way*, and *hard work* attain high t-scores but low MI scores, while pairs like *ultimate arbiter*, *immortal souls* and *tectonic plates* attain the reverse” (Durrant & Schmitt, 2009: 167; original italics). To tap into the sort of information that each measure provides, Durrant and Schmitt (2009) use these two association scores and analyse them independently of one another. Church et al. (1994) suggest another approach that involves intersecting the two measures and analysing word pairs that have high scores on both measures.

An alternative approach, recommended particularly for beginners and intermediate learners, involves the application of a modified version of the MI score called the Mutual Information cubed (MI3) score. The MI3 score rebalances the MI score in that it gives more weight to frequent words and less weight to infrequent words. The fact that the frequencies are cubed gives a much bigger boost to high-frequency word combinations than to low-frequency ones. As mentioned earlier, the current study made use of raw frequencies and the MI3 method (see Chapter 7, Section 7.5).

4.6. Phonological Short-term Memory and adult L2 collocation learning

Having discussed the approaches to defining and testing collocations, we shall now return to the concept central to this thesis, namely the relationship between Phonological Short-term Memory (PSTM) and L2 collocation learning. This section has direct links to Chapter 2, which presents PSTM within the context of the Working

Memory model, and Chapter 3 (Section 3.6), which outlines the link between PSTM and experimental, naturalistic and classroom-based vocabulary learning. The more extended account of collocations in the present chapter provides us with the context to explain what is novel and significant about our investigation of the link between PSTM and L2 collocation learning in adults. This section is a foundation for our subsequent interpretation of the patterns of results obtained in our longitudinal classroom-based study of the development of L2 collocational knowledge in relation to individual differences in PSTM among young adult subjects (see Chapter 7).

Out of the tests and techniques presented in Section 4.5, translation from L1 to L2 is the only one that has so far been used to link L2 collocation learning and PSTM capacity in A1 adult learners (Ellis & Sinclair, 1996). A laboratory-based experiment conducted by Ellis and Sinclair (1996) is the only study that has to date attempted to link PSTM and L2 collocational learning directly. To the best of our knowledge, there are no classroom studies, either cross-sectional or longitudinal, that have attempted to extend Ellis and Sinclair's findings to the classroom context and to different levels of L2 proficiency. Apart from describing the study by Ellis and Sinclair, this section also recapitulates the role of PSTM in L2 vocabulary learning at different proficiency levels, and presents differing viewpoints as to whether or not the role of PSTM in L2 collocation learning resembles that in vocabulary learning (Ellis & Sinclair, 1996; O'Brien et al., 2006). Indirect evidence for a possible link in the classroom context between PSTM and collocational knowledge in beginners and intermediate learners⁴⁵ is also presented below.

4.6.1. PSTM and collocation in adult L2 learners: experimental evidence

There is only one laboratory-based study that has addressed the issue of PSTM and L2 collocation learning in adult L2 learners (Ellis & Sinclair, 1996). The Ellis and Sinclair (1996) study set out to establish whether interfering with PSTM by means of articulatory suppression (see Chapter 2, Section 2.3.2 for a description of the technique) had an impact on the efficiency of learning new L2 utterances. It involved 87 English

⁴⁵ We will generally use the term "beginners" to refer to A1 and A2 learners and "intermediate students" to refer to B1 and B2 learners. Our own study deals with A2 (elementary) and B1 (pre-intermediate) students.

native speakers (mean age = 21), who were randomly assigned to three experimental conditions: silence, repetition and articulatory suppression. Subjects were instructed to learn L1 (English) translations of L2 (Welsh) phrases and had no prior knowledge of Welsh. As predicted, those subjects who were requested to repeat novel L2 expressions out loud (the repetition condition) were significantly better than those who were prevented from doing so (the articulatory suppression condition) or those who were instructed to remain silent (the silent condition). The articulatory suppression condition, which involves the use of a demanding secondary task that competes for the limited capacity of PSTM, was shown to produce the lowest number of correctly remembered L2 phrases, and the scores were significantly different from those obtained in the silent condition ($p < .05$) and in the repetition condition ($p < .001$). On the basis of the above Ellis and Sinclair conclude that PSTM is implicated in the acquisition of L2 phrases (1996: 245, see also Ellis, 1996 and Ellis & Schmidt, 1997). Ellis and Sinclair also suggest that the involvement of PSTM in L2 vocabulary learning resembles that for L2 collocation learning:

...collocations can simply be viewed as big words, and the role of working memory⁴⁶ in learning such structures is the same as that for words. Just as repetition aided the consolidation of Welsh vocabulary in the present experiment, so it did the long-term acquisition of Welsh phrases. (Ellis & Sinclair, 1996: 245)

The review of research on the role of PSTM in L1/L2 vocabulary learning in Chapter 3 (Section 3.6) indicates that PSTM involvement in vocabulary learning decreases as L1/L2 proficiency increases. Gathercole (2006a) points out that, with respect to the learning of individual words, more advanced L2 learners become less reliant on PSTM and more adept at using their mental lexicons to access phonological representations of close neighbours of the target item. The Ellis and Sinclair lab-based study deals with the initial stage of L2 learning, and it is not clear how their findings extend to other levels of L2 proficiency. A greater learning effort can be expected in the case of multi-word phrases when compared to the learning of single words (Verstraten, 1992: 28), and therefore it is plausible that PSTM may still play a role in the learning of collocations at post-elementary levels of proficiency, even though the

⁴⁶ Ellis and Sinclair are referring to PSTM here.

link between PSTM in L2 vocabulary learning is thought to diminish as the L2 proficiency increases.

4.6.2. PSTM and collocation in adult classroom L2 learners

Indirect evidence suggesting that there may be a significant link between PSTM and L2 collocation learning in the classroom context can be found in research on L2 grammatical learning in adult learners. The O'Brien et al. (2006) longitudinal study reveals that PSTM capacity, tapped by L1-based (English) Serial Nonword Recognition at Time 1, is significantly linked with L2 (Spanish) subordinate clause use at Time 2 in post-beginner⁴⁷ adult learners of Spanish ($r = .497, p < .01$). O'Brien et al. refer to the work of Romani (1994) and Speidel (1993) on individuals with impaired PSTM, whose syntactic deficits involve word order problems, to support their argument that PSTM plays an important role in the learning of word order in adult L2 learners (2006: 398). Apart from analysing data from their sample as a whole ($n = 43$), O'Brien et al. also present correlation patterns in the group of students who obtained higher scores on SAT II Spanish Test at Time 1 (the high ability group) and the group of students who obtained lower scores on the test (the low ability group). In their low ability group ($n = 23$) the Serial Nonword Recognition scores at Time 1 are significantly linked to free productive use of L2 vocabulary at Time 2 ($r = .423, p < .05$) and to L2 subordinate clause use at Time 2 ($r = .471, p < .05$). In their high ability group ($n = 20$), however, the Serial Nonword Recognition scores continue to be significantly linked with L2 subordinate clause use at Time 2 ($r = .532, p < .05$), but are no longer linked with gains in free productive use of L2 vocabulary at Time 2 ($r = .175, p > .05$). On the basis of the above O'Brien et al. conclude that at initial stages of language learning PSTM is mainly deployed for vocabulary learning, but that at later stages when vocabulary access is more automatic PSTM is redeployed for learning more complex structures (2006: 399). These results constitute preliminary evidence that the phonological learning mechanism may not be restricted to beginners only and that it may still operate on a multi-word level in more proficient users of the language.

⁴⁷ Regrettably, O'Brien et al. fail to report the proficiency level of their subjects in terms of the traditional distinction between beginners, intermediate and advanced students or other proficiency frameworks such as e.g. the CEFR (A1, A1, B1, B2, C1, C2). Admittedly, their subjects sat the SAT II Spanish test (used to place high school students into universities in the US), but O'Brien et al. do not indicate how the results obtained on the test reflect general levels of proficiency. O'Brien et al. report that participants all had at least two prior semesters of formal study of Spanish as L2.

A classroom-based cross-sectional study by Kormos & Sáfár (2008) on Hungarian pre-intermediate students of English (aged 15 and 16) points to a significant link between PSTM scores (tapped by L1-based Nonword Repetition) and the Use of English paper (consisting of a lexical and grammatical component), which suggest that PSTM may be active at post-elementary levels of L2 proficiency in adolescent learners. Separate figures for correlations between PSTM and the components of the Use of English test were not reported by the authors, however, which makes it difficult to compare the strength of the link between PSTM and L2 vocabulary and that between PSTM and L2 grammar.

The current study aims at exploring whether the findings regarding early L2 collocation learning in laboratory conditions and the indirect evidence from classroom learning studies can be extended to the classroom context (A2 and B1 proficiency levels) over a period of time. If the role of PSTM in L2 vocabulary learning is the same as that for L2 collocation learning, as Ellis and Sinclair (1996) suggest, then we would expect to obtain highly significant relationships between PSTM scores and L2 vocabulary/collocation scores in A2 learners, and non-significant (or significant but weaker) relationships between PSTM scores and L2 vocabulary/collocation scores in B1 learners. However, if the link between PSTM and L2 collocation scores remains strong in intermediate learners, then that would imply that PSTM is used for more complex structures at post-elementary levels of language learning, as suggested by O'Brien et al. (2006).

Chapter 5: Methodology and Piloting

5.1. Introduction

The present chapter comprises two main sections, one describing the design of the main study and the pilot study, and the other presenting the process of selecting, creating and validating our research instruments. Section 5.2.1 outlines the research questions posed in the main study (presented in greater depth in Chapter 1), lists the dependent and independent variables included in the main study, and presents the timeline for the administration of our research instruments. Section 5.2.2 focuses on the main objectives of the pilot study. These include, *inter alia*, obtaining reliability coefficients for our research instruments, and analysing the distribution of scores and item difficulty indices to ensure that our instruments distinguish between more and less able A2 and B1 Polish learners of English. Section 5.3 gives a detailed account of how each of the pilot study objectives was pursued, and what alterations had to be made to ensure that the instruments chosen for the main study were valid and reliable.

5.2. Research design

5.2.1. Design of the main study

Dependent and independent variables

The study identifies a number of dependent and independent variables (see Table 5.1 below). The *independent variables* include PSTM capacity with articulation (Serial Nonword Recall scores) and PSTM capacity without articulation (Serial Nonword Recognition scores). The *dependent variables* relate to knowledge and learning of L2 vocabulary and collocations, and include the following: L2 vocabulary learning outside the classroom, L2 vocabulary learning in class, L2 collocational learning outside the classroom, L2 collocational learning in class. Apart from measuring the rate of L2 vocabulary and collocational learning, the study also focuses on L2 vocabulary and collocation knowledge.

In addition, two potentially confounding variables were identified, namely the amount of exposure to L2 outside the classroom and productive knowledge of L1 (Polish) vocabulary. For detailed operational definitions of the constructs please consult Chapter 1, Section 1.2.

Table 5.1. Outline of dependent and independent variables

Type of variable	Variable name	Corresponding test scores
Independent variables	PSTM (with articulation) PSTM (without articulation)	Serial Nonword Recall scores Serial Nonword Recognition scores
Potentially confounding variables	Exposure to L2 outside the classroom Knowledge of L1 low-frequency vocabulary (measured productively)	number of hours per day recorded in students' journals L1 Vocabulary Test scores
Dependent variables	L2 collocation learning in class L2 collocation learning not related to class content L2 collocation knowledge not related to class content L2 vocabulary learning in class L2 vocabulary learning not related to class content L2 vocabulary knowledge not related to class content L2 proficiency level	A2 and B1 Progress Collocation Test scores change in A2 and B1 Gitsaki Collocation Test scores A2 and B1 Gitsaki Collocation Test scores at Time 1 and Time 2 A2 and B1 Progress Vocabulary Test scores change in A2 and B1 Vocabulary Levels Test scores A2 and B1 Vocabulary Levels Test scores at Time 1 and Time 2 A2 and B1 levels defined in relation to the Common European Framework of Reference (Council of Europe, 2001) and were measured by the OUP Placement Test (2001)

Note: L1 – Polish, L2 – English.

A2 and B1 Progress Vocabulary and Collocation Tests (Chapter 6, Section 6.3.6) probe knowledge of vocabulary and collocations that were taught in class. A2 and B1 Gitsaki Collocation Tests and A2 and B1 Vocabulary Levels Tests target words and collocations that were not directly taught in class but to most of which our subjects were likely to have been exposed outside the classroom. The selected sections of the Vocabulary Levels Test probe receptive and controlled productive knowledge in respect of the 1,000, 2,000 and 3,000 most frequent words (Section 5.3.3, and Chapter 6, Section 6.3.5). A2 and B1 Gitsaki Collocation Tests assess the controlled productive knowledge of selected grammatical and lexical collocations classified with accordance to the BBI typology (Section 5.3.2, and Chapter 6, Section 6.3.4). The frequency of occurrence of the collocations included in Gitsaki tests ranges from 7 to 16,090 occurrences in the British National Corpus.

The administration of the test battery

The test battery and times of test administration are presented in Table 5.2 below. The PSTM measures and the general vocabulary and collocation tests were administered before the commencement of the English language course (Time 1) and at the end of the course (Time 2). Gitsaki productive Collocation Test has one version only, and therefore the same version (with a different ordering of test items) was administered twice. As discussed in Section 5.3.3, the Vocabulary Levels Test comes in two parallel versions that are recommended for test/retest purposes (Laufer & Nation, 1999; Schmitt et al., 2001). However, not all pairs of different frequency sections turned out to be statistically equivalent when administered to A2 and B1 Polish learners of English during our piloting stage. In those few instances one of the available versions was chosen to be administered twice during the main experiment (see Section 5.3.3). The progress test probing knowledge of vocabulary and collocations taught in class was administered at Time 2.

Table 5.2. Testing battery and times of test administration

Times of administration	A2 students	B1 students
Time 1 (January 2008 –prior to the commencement of the language course)	A2 Vocabulary Levels Test (A): the receptive 1,000-word test (Nation, 2001b), the receptive 2,000-word test (Nation, 2001b modified by Schmitt et al., 2001; Schmitt, 2000), the productive 1,000-word test (Skrzypek, Section 5.3.3) & the 2,000-word productive test (Laufer & Nation, 1999) A2 Gitsaki productive Collocation Test (Gitsaki, 1999) L1 Vocabulary Test Serial Nonword Recall Serial Nonword Recognition	B1 Vocabulary Levels Test (A): the receptive 2,000-word & 3,000-word tests (Nation, 2001b modified by Schmitt et al., 2001; Schmitt, 2000), the productive 2,000-word & 3,000-word tests (Laufer & Nation, 1999) B1 Gitsaki productive Collocation Test L1 Vocabulary Test Serial Nonword Recall Serial Nonword Recognition
Time 2 (June 2008 – the end of the language course)	A2 Vocabulary Levels Test (B): (as above) ⁴⁸ A2 Gitsaki productive Collocation Test A2 Progress Vocabulary Test A2 Progress Collocation Test Serial Nonword Recall Serial Nonword Recognition	A2 Vocabulary Levels Test (B): (as above) B1 Gitsaki productive Collocation Test B1 Progress Vocabulary Test B1 Progress Collocation Test Serial Nonword Recall Serial Nonword Recognition

⁴⁸ Please note that not all pairs of levels tests turned out to be statistically equivalent when administered to A2 and B1 Polish learners of English during our piloting stage. In those few instances one of the available versions was chosen to be administered twice (see Section 5.2.3).

Research questions

Three of our research questions relate to the link between PSTM and L2 vocabulary learning in adults:

1. Does the relationship between PSTM (as tapped by the Serial Nonword Recall task and the Serial Nonword Recognition task) and L2 vocabulary learning (as tapped by the Vocabulary Levels Test scores at Time 2, the Vocabulary Levels Test gain scores and the Progress Vocabulary Test scores) decrease with an increased level of L2 proficiency in adults, as has been shown to happen in the case of child L1 learning?
2. Which component of PSTM is more closely linked to L2 vocabulary learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?
3. In the case of adults (as opposed to children who were the focus of previous research), what is the causal direction of the relationship between PSTM and L2 vocabulary learning at the A2 and B1 levels of L2 proficiency? Is it PSTM at Time 1 that contributes to L2 Vocabulary Levels Test scores at Time 2, or is it L2 Vocabulary Levels Test at Time 1 that contributes to the PSTM scores at Time 2?

The remaining three research questions relate to the link between PSTM and L2 collocation learning in adults, and they address the following aspects:

4. In the case of adults, is there a relationship between PSTM (as indexed by the Serial Nonword Recall and the Serial Nonword Recognition) and the learning of L2 collocations (as tapped by Gitsaki Collocation Test scores at Time 2, Gitsaki Collocation Test gain scores and the Progress Collocation Test scores) at the A2 and B1 levels of L2 proficiency? If so, does the relationship between PSTM and L2 collocation learning decrease with an increased level of L2 proficiency?

Assuming the relationship specified in Question 4 exists:

5. Which of the two components of PSTM, the articulatory component (tapped by the Serial Nonword Recall task) or the nonarticulatory storage component (tapped by the Serial Nonword Recognition task), is more closely linked to L2 vocabulary collocation in adults at the A2 and B1 levels of L2 proficiency?

6. What is the casual direction of the relationship between PSTM and L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency? Is it PSTM at Time 1 that contributes to Gitsaki Collocation Test scores at Time 2, or is it Gitsaki Collocation Test scores at Time 1 that contribute to the PSTM scores at Time 2?

The above research questions are discussed in greater depth in Chapter 1.

5.2.2. Objectives of the pilot study

A number of objectives were pursued at the piloting stage and these include the following:

1. creating nonwords for our PSTM measures and assessing their level of wordlikeness (Section 5.3.1),
2. obtaining reliability coefficients for our research instruments (Section 5.3),
3. analysing the distribution of scores from our research measures to ensure the obtained scores are spread well enough to distinguish between more and less able test-takers (Section 5.3),
4. investigating the subjects' performance in relation to English-Polish cognates included in the receptive and productive Versions 1 and 2 of the Vocabulary Levels Test (Section 5.3.3),
5. investigating the subjects' performance in relation to the T-F statements on the 1,000 Vocabulary Levels Test that draw on general knowledge (Section 5.3.3),
6. checking that the chosen frequency levels of the Vocabulary Levels Test are appropriate in terms of their difficulty for A2 and B1 learners of English (Section 5.3.3),
7. testing the equivalence of the amended versions of the Vocabulary Levels Test for Polish test-takers (Section 5.3.3),
8. establishing the level of correlation between Serial Nonword Recall scores and Serial Nonword Recognition scores (Section 5.3.1),
9. and establishing the level of correlation between vocabulary and collocation scores (as measured by the Vocabulary Levels Test and the Gitsaki test) in A2 and B1 Polish learners of English (Section 5.3.3).

All the above-listed objectives are explored in detail in Section 5.3 below.

5.3. Piloting of the instruments

5.3.1. Serial nonword tasks

The process of constructing experimental nonwords

The 65 nonwords created for use in this research project are pronounceable phonological sequences according to Polish phonotactics. Each item is a three-syllable CV⁴⁹-CCV-CV combination. The constituent syllables were selected from a Polish language dictionary (Polish Scientific Publishers PWN – CD-ROM edition) and a frequency dictionary of Polish (Kurcz et al., 1990). The position in which these syllables appear in the constructed nonwords reflects their position in real 3-syllable Polish words (i.e. initial, medial and final). For a list of syllables used in the final version of our Serial Nonword Recall and Serial Nonword Recognition tasks please consult Chapter 6, Sections 6.3.2 and 6.3.3.

The nonwords were recorded using high-quality recording equipment in a sound-proof cabin in the phonetics laboratory at Trinity College Dublin. The native speaker of Polish who provided the sound recordings was a teacher of the Polish language and had a particularly clear standard Polish pronunciation. The nonwords selected for the experiment had the same range of articulatory durations.

Wordlikeness ratings were obtained for the 65 nonwords in order to exclude those items that were too reminiscent of real Polish words. As discussed in Chapter 2, wordlikeness ratings reflect the degree to which novel syllable sequences resemble existing words. Such ratings are normally provided by native speakers of a language the nonwords are based on, as native speakers are believed to “have consistent intuitions about whether a sequence of speech sounds could be a word of their native language and how typical it would sound as a word” (Bailey & Hahn, 2001: 568). Bearing in mind that the perceived level of wordlikeness has an impact on performance on nonword tasks, the lower the degree of wordlikeness, the more limited the involvement of long-term memory and the purer the measure of PSTM (see e.g. Gathercole & Pickering, 1999).

⁴⁹ “C” stands for a consonant and “V” for a vowel.

Wordlikeness ratings for our nonwords were provided by 23 Polish adults studying English as their major at the University of Silesia in Poland (male/female = 8/12; age range = 20-26, $\mu = 21.3$, $SD = 1.26$). The students were instructed to listen to high-quality recordings of 65 nonwords in random order of presentation and rate them on a 5-point scale ranging from 1 (very unlikely to pass for a real Polish word) to 5 (very likely to pass for a Polish word). A mean wordlikeness rating for each nonword was then computed and 40 nonwords with the lowest ratings were selected (min/max = 1.3 / 2.35, $\mu = 1.988$, $SD = 0.259$). Wordlikeness ratings for the nonwords selected for Serial Nonword Recall and Serial Nonword Recognition are provided in Appendix J, Table A.15.

At the beginning of the rating session the students were asked to write down any real Polish word/s they felt a given nonword closely resembled. The nonword “cestura”, for example, was quite often associated with an actual Polish word “cenzura” meaning “censorship” (see Table 5.3). The subjects provided examples of similar-sounding Polish words for 92.5 % of the highly wordlike nonwords (rating ≥ 4) and for 15 % of those with middle ratings (≥ 2.5 and < 4). They did not provide any examples of similar-sounding Polish words for nonwords judged as minimally wordlike. The number of phonemes shared by highly/moderately wordlike nonwords and real Polish words that were said to bear some resemblance to the nonwords ranged from three to five (in bold in the table below).

Table 5.3. Piloting – examples of high-wordlike nonwords and similar-sounding Polish words

Examples of highly wordlike nonwords (rating ≥ 4)	Examples of similar-sounding Polish words	English translations of the Polish words
bafroda ^A cestura lukliwo miksoza moglena muklawy nekromu repluko zeklono zodrypa	broda cenzura szkliwo wiskoza migrena murawy nikomu replica klon grypa	beard censorship tooth enamel viscose migraine swards no-one (<i>dative</i>) replica clone flu

^A Phonemes shared by the nonwords and Polish words in bold.

In order to exclude such syllable combinations that may be relatively difficult to pronounce even for native speakers of Polish, the degree of difficulty of pronunciation

of each item was assessed on a scale ranging from “1-very easy to pronounce” to “5-very difficult to pronounce”. None of our 40 minimally wordlike nonwords was deemed as difficult or very difficult to pronounce by the Silesian University students participating in the piloting stage (range = 1.05 / 2.3, $\mu = 1.48$, $SD = 0.467$).

The 40 selected nonwords were subsequently divided into two pools of 20 items each, one for use in the Serial Nonword Recognition task and the other for the serial nonword repetition task. Bearing in mind the fact that similar-sounding memory items are remembered more poorly than those that sound different (e.g. Lewandowsky & Farrell, 2008), an effort was made, while assigning nonwords to the two separate pools, to ensure that the nonwords in each pool did not sound alike and that each constituent syllable was used only once within one pool of items.

The Serial Nonword Recall procedure involves repeating lists of nonwords immediately after the presentation of a given list in exactly the same order in which the items had been presented. In the Serial Nonword Recognition task, on the other hand, no articulation is required. Instead the subjects are asked to judge whether nonwords within pairs of nonword lists are presented in the same order or not. Our Serial Nonword Recall task consists of three sets of nonword lists, each made up of 15 lists of the same length. Set one, two and three contain 2-item, 3-item and 4-item lists respectively (see Appendix C). Our Serial Nonword Recognition task is comprised of 30 pairs of nonword lists with 10 pairs of nonword sequences at each of three list lengths, that is at 4-item, 5-item and 6-item lengths (see Appendix D). In the serial recall task one point is scored for a correctly repeated list. In the serial recognition task one point is scored for a correctly recognised set of lists as either the same or different. The maximum score on the recall and recognition task is 45 and 30 respectively.

Normality and distribution of scores

Serial Nonword Recall and recognition tasks were piloted on a group of 20 young Polish nationals resident in Ireland (male/female = 6 / 14; age range = 25-34, $\mu = 29$, $SD = 2.85$; length of residence = 18.5 months, $SD = 9.19$). The subjects were recruited over a period of three months via free advertisements placed in *Polska Gazeta*, a popular Polish newspaper. No payment was offered for participating in the current study. The Poles were tested on a one-to-one basis in a language laboratory in

the Centre for Language and Communication Studies (CLCS), Trinity College Dublin. The order of administration of the recall and recognition tasks was counterbalanced. 10 of our subjects, randomly chosen, completed the recall task first, while the remaining 10 were asked to complete the recognition task first. The results were then entered into a statistical package, SPSS 16.0⁵⁰, to analyse the distribution of scores and the degree of correlation between the two measures.

Each of our subjects was recorded during Serial Nonword Recall in order to calculate intra-rater reliability. We anticipated that making voice recordings might increase the participants' levels of self-awareness and lower their performance level. With this concern in mind, we availed of CLCS language lab facilities, as they allow the making of voice recordings in a fairly unobtrusive manner. Each test taker sat comfortably in front of a computer screen, where task instructions were displayed, and listened to recordings through high quality headphones with inbuilt microphones. In order to obtain more information on how consistently the rater – that is the author of this thesis – assessed the students' responses, the recall task was analysed at two levels – at the list level and the item level. One point was given for each correctly repeated list that did not contain any mistakes. Additionally, the rater marked correctly pronounced nonwords and their order on a designated webpage developed for the piloting of this instrument. The second rating was based on the analysis of students' recordings and took place approximately one week after the first rating. The obtained Cronbach's Alpha coefficient of .999 is very high. The rater scored one list as incorrectly repeated. The analysis of the scoring of individual nonwords revealed that the rater had misheard the way one nonword was pronounced.

Serial Nonword Recall scores range from 6 to 27, while serial recognition scores range from 9 to 23 (see Table 5.4). The values of skewness and kurtosis⁵¹ are very close to zero, as would be expected in normally distributed data. The actual values of skewness and kurtosis are not informative in themselves, though, until they are converted to a z-score. When transformed to a z-score, the values of skewness and kurtosis for both distributions are non-significant, which confirms the assumption that both sets of scores are normally distributed. Additionally, the Kolmogorov-Smirnov

⁵⁰ SPSS 16.0 is available from <http://www.spss.com/ie>.

⁵¹ Skewness characterises the degree of asymmetry of a distribution around the mean. Kurtosis characterises the degree of flatness or peakedness of the distribution as compared to the normal distribution.

test (Table 5.4 below) and normal probability plots (Appendix J, Figures A.1 and A.2) indicate that neither of the two distributions significantly deviates from normality.

Table 5.4. Piloting – distribution of scores in serial nonword tasks (n=20)

Serial Nonword Recall	k	45
	μ	15.25
	min /max	6/27
	SD	5.45
	Kolmogorov-Smirnov test	0.2, ns
	Skewness	0.076, ns
	Kurtosis	-0.508, ns
	Spearman-Brown split-half reliability (for unequal length)	0.95
Serial Nonword Recognition	k	30
	μ	16.1
	min /max	9/23
	SD	4.04
	Kolmogorov-Smirnov test	0.2, ns
	Skewness	-0.123, ns
	Kurtosis	-1.038, ns
	Guttman split-half reliability	0.86

There is a correlation between Serial Nonword Recall and recognition scores ($r = 0.675$, $p < .01$), and its strength corresponds to results obtained in other studies employing these two procedures (e.g. Gathercole et al., 1999a). The Pearson product-moment correlation coefficient increases to $r = 0.781$ when it is corrected for attenuation using Equation 5.1 below:

$$r = \frac{r_{XY}}{\sqrt{r_{XX'} r_{YY'}}$$

Equation 5.1. Correcting correlation coefficients for attenuation

- r – the estimated correlation between the ‘true scores’ for X and Y
- r_{XY} – the obtained correlation coefficient
- $r_{XX'}$ – the reliability coefficient for the scores of variable X (Table 5.2)
- $r_{YY'}$ – the reliability coefficient for the scores of variable Y (Table 5.2)

To assess the reliability of the serial nonword tasks split-half reliability estimates were calculated. In order to estimate split-half reliability the test scores are normally split into two halves. As a result, each test taker has two scores, one for each half of the test. To ensure that the two halves are equivalent in what they measure, a rational split based on some criterion is recommended in some cases rather than a random split (Bachman, 2004: 161). We adopted a combination of a rational and random split with list lengths being our criterion for the split. The test scores were

divided randomly at each list length rather than across the whole test. In Serial Nonword Recognition task, at each list length 10 scores were randomly split into two halves and assigned to each half-test scores. To estimate internal consistency reliability on the basis of the two halves, the Guttman split-half reliability estimate was calculated with the formula in Equation 5.2:

$$r_{xx} = 2 \left(1 - \frac{s_{h1}^2 + s_{h2}^2}{s_x^2} \right)$$

Equation 5.2. Calculating Guttman split-half reliability estimate

r_{xx} – the reliability coefficient

s_{h1}^2 – the estimator of the population variance of one set of half-test scores

s_{h2}^2 – the estimator of the population variance of the other set of half-test scores

s_x^2 – the estimator of the population variance of the total test scores

In Serial Nonword Recall the number of items was odd ($k = 45$), which resulted in two halves of unequal length. To correct the correlation for length the Spearman-Brown split-half reliability estimate was calculated. The formula is given in Equation 5.3:

$$r_{xx} = \frac{2r_{hh}}{1 + r_{hh}}$$

Equation 5.3. Calculating the Spearman-Brown correction for length

r_{xx} – the reliability coefficient

r_{hh} – the correlation between the two halves

The piloting of Serial Nonword Recall shows that the instrument is highly reliable. The obtained split-half reliability estimate is 0.95. The recognition task has an adequate split-half reliability estimate of 0.86. Reliability coefficients over .90 indicate very good reliability, while coefficients between .80 and .89 represent acceptable reliability (Vierra & Pollock, 1992). Tests with the latter level of reliability are suitable for use in research. They would yield more reliable results, however, if they were re-administered a small number of times and then averaged with scores of similar reliability. It has to be kept in mind that administering one test a number of times is very impractical and rarely feasible. Considering the obtained reliability coefficient for our recognition task, we decided to administer the task twice at the beginning of the main experiment and calculate its test-retest reliability.

The piloting of Serial Nonword Recognition has alerted us to a potential problem with this measure when employed to tap PSTM capacity in adults. Serial Nonword Recognition has mainly been used in research on PSTM and child L1 learning (e.g. Gathercole et al., 1999b), and has not yet been extensively explored in studies of adult L2 learning (O'Brien et al., 2006). Adults are known to have at their disposal test-taking strategies and rehearsal strategies that young children do not possess. Our piloting shows that some adults attempt to employ such strategies to obtain a higher score on the Serial Nonword Recognition task. Two of our subjects, by their own admission, tried to lower the memory burden by holding only each initial syllable in their memory instead of the whole word. To counteract that problem and to try to ensure that our testees attempted to remember the whole nonwords, the subjects in our main experiment were informed that the testing session would also involve a cued-recall task to assess how many of the presented nonwords they could retrieve from memory. Having completed the cued-recall task, they were also asked for feedback on the mechanisms they used to remember the nonword lists.

5.3.2. Gitsaki Collocation Tests

Gitsaki (1999) developed collocation tests for three English proficiency levels, i.e. for beginners, intermediate and upper-intermediate students. For a number of reasons, we decided to include her test for beginners (henceforth referred to as the A2 test) and the test for intermediate students (the B1 test) in our test battery as a measure of general collocational knowledge. First, when administered to a relatively large sample of English learners (91 beginners, 94 intermediate and 90 upper-intermediate Greek learners of English), the tests revealed significant differences in controlled productive knowledge of collocations at these three levels of proficiency. Second, the tests contain a number of different types of both grammatical and lexical collocations. Grammatical collocations, as classified by the BBI typology (Benson et al., 1986), are those patterns that consist of an open and a closed class word (e.g. adjective + preposition), while lexical collocations comprise open class words (e.g. verb + noun). Finally, the collocations the tests contain have a wide range of frequencies of occurrence in the BNC corpus.

In the foreword to the publication describing the study (Gitsaki, 1999) Professor Ali Aghbar expresses the view that “Gitsaki’s research methodology and, in particular,

her elicitation instruments (...) will prove very valuable to those who want to pursue a similar research design.” Despite having administered the tests to a considerable number of test-takers, though, Gitsaki failed to report the tests’ reliability coefficients and to calculate item difficulty indices. Therefore, it is essential that her fill-in-the-blank tests be properly validated before they can be re-used in any other learning context. To assess the tests’ reliability and suitability for A2 and B1 learners in the present study, their test-retest reliability and item difficulty coefficients were calculated.

Minor alterations to Gitsaki Collocation Tests

A number of minor adjustments had to be made prior the first administration of Gitsaki Collocation Tests. A few sentences on the A2 test that sounded very much like classroom instructions for a role-play were re-phrased (Sentences 18, 25, 30 and 33; see Appendix E). One test item had to be re-formulated for contextual purposes – “This is the story” was replaced with “This is a very long story” to make it more coherent with the sentence that follows (Sentence 39). The A2 test in its original form contained quite a few references to Greece. Most of these were replaced by references to places and names more familiar to Polish students of English. Additionally, Sentence 43 (“Every morning the school bus comes _____ 7.06”) was replaced by “The school _____ comes at 7.06 every morning and takes us to school”, because arriving at a reliable frequency count of “at (hour)” in the BNC turned out to be rather tricky even with the help of the Simple Query Syntax Guide for BNCweb available online. Finally, in Sentence 48 the noun “book” was used instead of the noun “musical”, as we felt our subjects were likely to be more familiar with the former.

A number of cosmetic changes were introduced into the B1 test as well. In Sentence 2 the word “only” was added. Sentence 11 (“He failed _____ (in) his university examinations”) was replaced with “Doctors fail _____ (in) their duty if they do not warn their patients of the dangers.” The original sentence remains grammatically correct when the preposition “in” is omitted. Sentence 28 (“The referee _____ a penalty and the game stopped”) was replaced by “He looks terrible. I think he’s _____ the penalty for all those late nights.” We felt that the original sentence might be a bit problematic for those test-takers who are not too familiar with the rules of football/soccer. In Sentence 32 “...and I can’t find it” was added. In Sentence 34 an

English name John was used. The headline “Mother Finds Missing Son _____ Accident” in Sentence 58 was replaced by “mother had found her missing son _____ accident.” We felt this would be easier for B1 students to understand. For the same reason, in Sentence 64 (“The accusations _____ Peter made the judge put him in prison”) the latter part of the sentence (“...made the judge put him in prison”) was replaced by “were very serious, and he was put in prison.”

Normality and distribution of scores

The A2 and B1 Gitsaki Collocation Tests were piloted on 15 A2 learners (male/female = 6/9; age range = 22-29, $\mu = 23.93$, $SD = 1.71$) and 15 B1 learners (male/female = 5/10; age range = 23-26, $\mu = 23.87$, $SD = 0.99$) respectively. The subjects were students of English as a minor at the Higher School of Marketing in Silesia, Poland. In addition to Gitsaki Collocation Tests they were also asked to complete relevant sections of the Vocabulary Levels Test so that a comparison could be drawn between their lexical and collocational knowledge. The Vocabulary Levels Test results are presented in depth in the part of Section 5.3.3 that analyses the behaviour of the amended versions of the test. The collocation and vocabulary tests were administered in a pen and paper format, as we did not have access to the necessary computer facilities in the college to administer the test online. In order to avoid cheating the test items were re-arranged to create two versions for each test.

The scores obtained on the A2 and B1 Gitsaki Collocation Tests are normally distributed. The values of skewness and kurtosis are very close to zero and are all non-significant. The Kolmogorov-Smirnov coefficients are non-significant as well. Boxplots and normal probability plots (Appendix J, Figures A.5 to A.8) also indicate that the two distributions do not significantly deviate from normality. Gitsaki Collocation Tests were re-administered after a two-week period in order to calculate the tests' reliability measures. Table 5.5 contains data from the first administration of the collocation tests.

Table 5.5. Piloting – distribution of scores in Gitsaki Collocation Tests (n=15)

A2 Collocation Test	μ	37.06 %
	min /max	22 / 54 %
	SD	9.88 %
	Kolmogorov-Smirnov test	0.158, ns
	Skewness	-0.193, ns
	Kurtosis	-0.937, ns
	test-retest reliability	0.91
B1 Collocation Test	μ	49.74 %
	min /max	30.8 / 64.6 %
	SD	9.02 %
	Kolmogorov-Smirnov test	0.078, ns
	Skewness	-0.352, ns
	Kurtosis	-0.01, ns
	test-retest reliability	0.95

Note: The results are reported in percentages as the total number of items per test is different at each level ($k_{A2}=50$ and $k_{B1}=65$).

The test-retest reliability coefficients for the A2 and the B1 Gitsaki Collocation Tests were 0.91 and 0.95 respectively. These coefficient values indicate that test-takers obtain very similar scores when the tests are re-administered within a short period of time. The minimum score obtained on the A2 and B1 tests were 22 and 30.8 per cent respectively. The maximum scores were 54 and 64.6 per cent respectively.

Item difficulty index

In order to assess the difficulty of the test an item analysis was carried out. The item difficulty index is the most useful, and possibly most frequently reported, type of item analysis statistic. Calculating the item difficulty index is an important part of a test piloting process in that it can help to identify potential problems with individual test items. The shape of the test score distribution, and the overall consistency and reliability of the test are directly affected by the levels of difficulty of individual test items. In norm-referenced testing (NRT), where the focus is on ranking the entire set of test-takers to make comparisons of their performances relative to one another, the scores are expected to be spread out. A test that had too high or too low item difficulty, with the scores bunched up together, would not allow any meaningful comparisons.

The analysis of item difficulty involves calculating the proportion of test takers who answer the item correctly. For small groups of subjects we can calculate the item difficulty index using the simple formula below (Bachman, 2004: 125):

$$p_i = \frac{R_T}{N}$$

Equation 5.4. Calculating item difficulty for small groups of subjects

R_T – the total number of test takers who got the item correct

N – the total number of test takers

The values of the item difficulty statistic range between 0.0 and 1.0, with a higher value indicating that a larger proportion of test-takers responded correctly to that particular item. Thus, an item with a difficulty level of 0.9 would be considered relatively easy, while an item with a p-value of 0.1 would be rather difficult. Items that have either a very high or very low p-value do not discriminate well between examinees with different levels of knowledge or skill. The average p-value for many items in NRT would therefore be expected to range between 0.4 and 0.6 for the test scores to be spread out well enough to provide a reliable ranking of the test-takers performance relative to one another.

The average item difficulty coefficients for the A2 and B1 test are 0.37 and 0.49 respectively. These values indicate that the A2 test is on average more difficult for the A2 test-takers than the B1 test is for the B1 test-takers. The analysis of item difficulty reveals that the A2 test has a higher number of test items with very few (or no) correct responses, while the B1 test has a higher number of items with over 90 per cent of correct answers (see Table 5.6 for some examples). One of the A2 items with the item difficulty coefficient of 0.0 that particularly attracted our attention is the phrase “points of the compass”. Given its very low number of occurrences in EFL teaching materials and in the BNC corpus (see Table A.1 in Appendix E), we feared that even B1 students would not be very likely to know this phrase productively. Therefore, we decided to replace it with another “noun + (of) + noun” phrase – “a slice of bread” (L5-type collocation according to the BBI typology, 140 occurrences in the BNC).

Table 5.6. Piloting – collocations with the highest and lowest item difficulty coefficients

Proficiency level	Collocations	No on the test	Item difficulty	Collocations	No on the test	Item difficulty
A2	do homework	4	0.93	ask about	33	0.00
	at the moment	26	0.87	complain about	21	0.00
	get up	28	0.87	crowded with	5	0.00
				make a		
	look for	34	0.80	suggestion	18	0.00
	hold on	31	0.73	piece of advice	35	0.00
				points of the	9	0.00
	on Sunday	19	0.73	compass		
	phone number	20	0.73	time flies	2	0.00
	school bus	43	0.73	famous for	38	0.07
	be on (of a programme)	42	0.67	lie down	49	0.07
	ask for	25	0.60	take sth out (of a bag)	36	0.07
B1	full of	10	1.00	directions on	35	0.00
				hold an		
	in fact	2	1.00	exhibition	7	0.00
				think highly		
	in the future	40	1.00	(of)	4	0.00
				attitude		
	look at	13	1.00	towards	56	0.07
	believe in	36	0.93	brake hard	17	0.07
				convince	3	0.07
	by the way	50	0.93	somebody of		
	find out	12	0.93	fail in	11	0.07
	in a panic	18	0.93	go out	20	0.07
	very well	16	0.93	take shorthand	24	0.07
	ask for	48	0.87	on the contrary	41	0.13

Note: The words in bold are the missing items in the test.

We acknowledge that there is a difference between average item difficulty on the two tests, but do not consider this a major problem. What is more important from our perspective is the fact that when the relevant tests were administered to the A2 and B1 students, the scores obtained should be spread out enough to distinguish between higher and lower ability students.

5.3.3. Vocabulary Levels Test

This section looks at the issue of whether parallel versions of the Vocabulary Levels Test are equivalent when administered to Polish learners of English. Our motivation for the choice of frequency levels for the A2 group (1,000- and 2,000-word frequency tests) and the B1 group (2,000- and 3,000-word frequency tests) is also outlined. A number of minor alterations introduced for linguistic and contextual reasons are presented. Finally, the development and piloting of the productive 1,000 Vocabulary Levels Test is discussed.

The receptive Vocabulary Levels Test (e.g. Nation, 2001b; Schmitt, 2000; Schmitt et al., 2001) comprises separate tests for the 1,000, 2,000, 3,000, 5,000, 10,000-word levels and the University word level (see Chapter 6, Section 6.3.5 for more detail). The test has been used by a considerable number of researchers in cross-sectional and longitudinal studies of vocabulary size in second language learners (e.g. Cameron, 2002; Cobb, 2000; Cobb & Horst, 2004; Fan, 2000; Laufer, 1998; Laufer & Paribakht, 1998; Xing, 2007; Xing & Fulcher, 2007), with the latter type being more frequent than the former. The 1,000-word level (Nation, 2001b) selected for use in this study comprises 39 True or False test items. All the other receptive tests have a multiple-choice format, and each probes the receptive knowledge of 30 words from the relevant frequency band. The test-takers are provided with 30 definitions divided into 10 sets of three definitions and six words each, and are expected to match each definition with one of the six words provided within the same set.

The productive Vocabulary Levels Test (Laufer & Nation, 1999) contains separate tests for frequency levels from the 2,000-word level to the University word level as well. Each of the tests consists of 18 fill-in-the-blank sentences, with a variable number of initial letters provided to ensure that there is only one correct word that fits the gap. For the purpose of our own project a productive 1,000-word level test was developed. The process of developing and piloting of this test is presented in detail later in this chapter.

Parallel versions of the Vocabulary Levels Test

A number of versions of the receptive Vocabulary Levels Tests have been created and validated (e.g. Beglar, 2009; Beglar & Hunt, 1999; Nation, 1993; Read, 1988; Schmitt, 2000; Schmitt et al., 2001). The four original versions, A-D, had 18-items per frequency level, and were shown to yield highly consistent results with coefficients ranging from .91 to .94 (Read, 1988). One of the versions used in Japan's Temple University as part of their test battery for English proficiency assessment and placement purposes has been found reliable at .95 and .97 (quoted in Beglar & Hunt, 1999). The 18-item versions have been extended to include 30 items per level, and have been subject to stringent statistical analyses (see e.g. Beglar & Hunt, 1999, for a validation of Versions A and B of the 2,000-word level). Additional 30-item per level versions have been developed as well (e.g. Schmitt, 2000), and have been

recommended for test-retest purposes in longitudinal studies (e.g. Schmitt et al., 2001). The most recent Vocabulary Size Test consists of 10 items for each consecutive level from 1,000 to 14,000 (see Beglar, 2009, for a detailed description and validation).

The two versions of the receptive 2,000 and 3,000 Vocabulary Levels Test selected for the current study include the 2000 Schmitt test – henceforth referred to as Version 1 – and the 2001 Nation test modified by Schmitt et al. (2001) – henceforth referred to as Version 2. The Nation parallel 1,000-word level tests (Nation, 2001b) were chosen to prompt the knowledge of the first 1,000 most frequent words. As for the productive Vocabulary Levels Test, our test battery includes Versions 1 and 2 of the 2,000 and 3,000-word frequency tests from Laufer and Nation (1999). The two versions of the receptive and productive Vocabulary Levels Test for the 2,000 and 3,000-word frequency bands have undergone stringent statistical analyses and are recommended for test-retest purposes (Laufer & Nation, 1999; Schmitt et al., 2001). Schmitt et al. (2001) report that when Version 1 and Version 2 of the receptive test are subjected to a Principal Component Analysis, each version loads with the same degree of magnitude on the same factor (.980 and .981 for 2,000 and 3,000-word versions respectively). The tests' means and variances are not statistically different for either Romance speakers or non-Romance speakers. As for the receptive Vocabulary Levels Test, Laufer and Nation (1999), created and validated four versions of the receptive test. They explored the behaviour of individual items on these tests and conflated them into two versions – Version 1 and Version 2. The authors say that – on the basis of the analysis of all items from the original four versions – Versions 1 and 2 can be considered equivalent in terms of means and variances.

It was deemed necessary in the present study to establish the equivalence of the parallel forms of the Vocabulary Levels Test when administered to Polish learners of English. Our line of reasoning is partially motivated by the fact that the information provided regarding the validation process of the various sets of tests varies in terms of its depth and quality. While Schmitt et al. (2001) and Laufer and Nation (1999) report their validation process in minute detail, Nation (1993) provides only very general information to support his claim that the two versions of the 1,000-word level test are equivalent. Another reason is the fact that the parallel versions have been validated on students of linguistic backgrounds other than Polish and therefore may not necessarily

be equivalent for Polish learners of English (see e.g. Slovakian subjects' results in Schmitt et al., 2001: 78). Words are known to have varying levels of difficulty for learners of different first languages even if they are sampled from the same frequency level (Laufer, 1997). Additionally, certain minor alterations (discussed later in this chapter) had to be introduced to the tests in order to increase their suitability for Polish test-takers. In the light of the above, we felt that it was essential that the behaviour of the two (slightly amended) versions be analysed prior to the main study.

The choice of frequency levels

The lexical aspect of the current study is intended to focus on the development of the A2 students' vocabulary knowledge within the 1,000 and 2,000-word frequency levels and the B1 students' knowledge of the 2,000 and 3,000-word frequency levels. Our motivation for the choice of these frequency levels is twofold. First, vocabulary lists linked to the CEFR descriptors, that is the Waystage A2 vocabulary list (van Ek & Trim, 1990b) and the Threshold B1 vocabulary list (van Ek & Trim, 1990a), indicate that A2 and B1 learners would be generally expected to be familiar with approximately 1,000 and 2,200 words respectively (Milton, 2009: 69). Second, considering the number of tests our subjects would be requested to complete during the main experiment, homing in on the most relevant frequency bands rather than administering the whole test battery appeared a reasonable compromise. Admittedly, the administration of the 1,000, 2,000, 3,000 and 5,000-word tests to both proficiency groups would provide more detailed information about the students' lexical development, but it is not uncommon for researchers to focus on one or two selected frequency levels (e.g. Beglar & Hunt, 1999).

Vocabulary size estimates linked to the CEFR descriptors are known to vary across studies depending on the second language under scrutiny and the country where the testing is taking place (see e.g. Meara & Milton, 2003; Milton, 2006b, reanalysed in Milton, 2009: 31; Milton & Alexiou, 2009). Meara and Milton's calculations of EFL vocabulary sizes (2003: 8) are, for example, slightly higher than those based on the Waystage and the Threshold vocabulary lists. According to their findings learners progressing from the A1 to the A2 level should be familiar with 1,500 words, while those progressing to the B1 level can be expected to know 2,500 words (see Chapter 3, Section 3.5). In the light of the above, it was necessary to establish that the selected

levels were indeed the most appropriate for our target proficiency groups, A2 and B1 Polish learners of English. Due the ceiling effect, an insufficiently challenging test would fail to provide adequate information about students' lexical gains over time. Therefore, during the piloting of the parallel versions of the Vocabulary Levels Test (later in this chapter) attention was paid to the level of test difficulty for A2 and B1 students.

Minor alterations to the Vocabulary Levels Test

For linguistic and contextual reasons a few minor alterations were made to the 1,000-word test (Appendix F). In sentence 20, Version 1 – “Silver costs a lot of money” – the word “silver” was replaced by the word “gold” (both words are within the 2,000 frequency word band). Silver is a relatively cheap commodity in Poland, while in Ireland it is rather expensive. Gold, on the other hand, is expensive in both countries. The picture of the girl provided alongside sentence 22 was replaced by another one, as the original picture could easily be taken for an image of a boy with longer hair. Having consulted a number of native speakers of English we also decided to rephrase sentences 3 and 15 in Version 2 to make them sound more natural. Sentence 3 was altered from “When someone asks ‘What are you called?’ you should say your name” to “When someone asks ‘What is your dog called?’ you should say your dog's name.” In sentence 15 – “People follow orders of a chief” – a definite article “the” was inserted in front of the noun “orders”. In sentence 28 the Polish name for Rome (“Rzym”) was added in brackets, as it was feared that not every student may be familiar with the English name. Two other city names used in this test, New York and London, are so widely known that we did not feel the need to provide their Polish equivalents.

We decided to look more closely at how test-takers respond to items that are cognates, that is have a similar/identical spelling in English and Polish. Considering the fact that the Vocabulary Levels Test does not probe the knowledge of all the lexical items within specified frequency bands, but only a small selection of items representing each band, the presence of a considerable number of cognates on the test could significantly change the test results. The degree of cognateness, along with such factors as word's length and its pronunciation, concreteness, imageability and part of speech, is known to influence word difficulty (Cobb, 2000; Gairns & Redman, 1986; Horst & Meara, 1998; Laufer, 1997; Milton, 2009). Owing to the fact that certain lexical items

occur in identical or similar-looking written forms in two or more languages, English words such as “hotel”, “taxi” or “pub” can be easily understood in their written form by speakers of some other European languages, e.g. French, Spanish or Italian (Gairns & Redman, 1986: 67). Consequently, test-takers can take advantage of the similarities between cognates’ written forms and provide a correct answer without having ever been exposed to that particular foreign target word.

The fact that a lexical item is a cognate does not necessarily mean that it has to be removed from frequency-based vocabulary tests when these are administered to speakers of a common native language, though. First, learners’ reactions to cognates can be highly idiosyncratic. As Milton (2009: 36) points out in relation to Limper’s study (briefly described in Tharp, 1934: 129-130), some learners view certain pairs of words as cognates, while others see them as completely unrelated. Likewise, a derivation of a word can be perceived as a cognate by some but not others. Second, cognates are a normal part of the make-up of different frequency bands. Since item selection from relevant bands is random, frequency-based tests of vocabulary knowledge, such as the Vocabulary Levels Test or the X-Lex Vocabulary test (Meara & Milton, 2003), are not unlikely to feature some of them. After careful consideration, therefore, we decided to calculate an item difficulty index for cognates on the Vocabulary Levels Test and to conduct short post-test interviews in order to assess their suitability for Polish test-takers.

Versions 1 and 2 of the 2,000 and 3,000-word level tests contain 15 English-Polish cognates, and these include “apartment” / “*apartament*”, “copy” / “*kopia*”, “education” / “*edukacja*”, “import” / “*import*”, “motor” / “*motor*”, “museum” / “*muzeum*”, “nerve” / “*nerw*”, “normal” / “*normalny*”, “original / *oryginalny*”, “private” / “*prywatny*”, “reform” / “*reforma*”, “structure” / “*struktura*”, “temperature” / “*temperatura*” and “wine” / “*wino*”. Some of the cognates are tested both receptively and productively, while others are tested in one mode only.

The meanings of these pairs of cognates overlap to a large extent with the exception of the pair “apartment” / “*apartament*”. The word “apartment” primarily means a place for people to live in that is part of a larger building and consists of a set of rooms. This very meaning is tested in the Vocabulary Levels Test. The Polish word “*apartament*”, on the other hand, denotes a set of very expensive hotel rooms (“suite”

or “bridal suite”), but it is not normally used to refer to a set of rooms in an apartment block, unless they are high-standard rooms occupied by a very important person such as e.g. a president. The English word “apartment” can take the latter meaning as well.

With a view to analysing the item difficulty of the cognates relevant test items were administered to 15 A2 Polish learners of English (male/female = 6 / 9; age range = 29 / 35, $\mu = 32.67$, $SD = 2.41$; length of residence in Ireland = 5.5 months, $SD = 2.134$). The number of cognates present in the receptive and productive versions was 14 and nine respectively. In receptive tests six of the cognates are included in Version 1, while the other eight are part of Version 2. In productive tests there are three cognates in Version 1 and five cognates in Version 2. Since it was the item difficulty of the cognates that was our primary focus, it was decided that instead of administering both versions of the test a number of clusters would be selected from receptive Versions 1 and 2 and combined into one receptive test ($k = 40$). A number of sentences were selected from productive Versions 1 and 2 as well to create one receptive test ($k = 15$).

The analysis of the students’ answers revealed that of all the cognates “motor” and “apartment” had the highest difficulty coefficients. On the receptive test both items obtained the item difficulty coefficient of 1.0, while on the productive test “motor” and “apartment” obtained the coefficients of 0.74 and 0.67 respectively (see Table 5.7 below). “Education” and “wine” were the third and fourth item in terms of their ease, with difficulty coefficients of 0.93 and 0.8 on the receptive test. Unusually high item difficulty coefficients indicate that an overwhelming majority of test-takers provide correct answers to the test items in question. A high item difficulty coefficient does not necessarily render a test item unsuitable, as some words within a given frequency band are bound to be easier for students than others (Milton, 2009; Schmitt et al., 2001). However, considering the low English proficiency of our subjects some of these correct answers could have been educated guesses. Other items had lower difficulty coefficients, namely 0.13 – 0.6⁵². Item difficulty coefficients of up to and including 0.6 were deemed acceptable.

⁵² Item difficulty coefficients were in almost all cases higher on the receptive test than on the productive one. One item that did not follow that pattern was word “normal”. Interestingly, it had a lower receptive coefficient value as compared to its productive value. We suspected that a slight difference in the meaning in the productive and receptive contexts might have contributed to higher productive figures of that word. On the productive test the adjective “normal” translates into Polish as “normalny”. However, on the receptive test the same adjective is used as a synonym of the word “usual” (as in e.g. “normal

Table 5.7. Piloting – item difficulty coefficients of cognates from the 2,000 and 3,000 Vocabulary Levels Test (n=15, level=A2)

English / Polish cognates	Part of speech	Receptive tests	Productive tests	Item difficulty	
				Receptive	Productive
apartment / <i>apartament</i>	noun	3,000-word level, Version 1, items 13	3,000-word level, Version 2, item 1	1.00	0.67
copy / <i>kopia</i>	noun	2,000-word level, Version 2, item 3	2,000-word level, Version 2, item 9	0.60	0.40
education / <i>edukacja</i>	noun	2,000-word level, Version 1, item 7	—	0.93	—
import / <i>import</i>	noun	3,000-word level, Version 2, item 8	3,000-word level, Version 1, item 5	0.4	0.27
motor / <i>motor</i>	noun	2,000-word level, Version 2, item 2	2,000-word level, Version 2, item 8	1.00	0.74
museum / <i>museum</i>	noun	3,000-word level, Version 2, item 3	—	0.60	—
nerve / <i>nerw</i>	noun	3,000-word level, Version 2, item 9	3,000-word level, Version 1, item 8	0.33	0.26
normal / <i>normalny</i>	adj	3,000-word level, Version 2, item 28	3,000-word level, Version 2, item 16	0.34	0.47
original / <i>oryginalny</i>	adj	2,000-word level, Version 1, item 25	—	0.26	—
popular / <i>popularny</i>	adj	2,000-word level, Version 2, item 30	2,000-word level, Version 1, item 18	0.60	0.47
private / <i>prywatny</i>	adj	2,000-word level, Version 1, item 26	—	0.60	—
reform / <i>reforma</i>	noun	—	3,000-word level, Version 1, item 3	—	0.13
structure / <i>struktura</i>	noun	3,000-word level, Version 1, item 6	3,000-word level, Version 2, item 6	0.60	0.26
temperature / <i>temperatura</i>	noun	2,000-word level, Version 1, item 4	—	0.53	—
wine / <i>wino</i>	noun	2,000-word level, Version 2, item 10	—	0.80	—

Note: In the receptive and productive versions of the 2,000 and 3,000-word level tests some lexical items are tested both receptively and productively, while others are tested in one mode only. For more information please consult the section on comparing the behaviour of the amended versions of the Vocabulary Levels Test later in this chapter.

Item difficulty coefficients for the words “motor” and “apartment” as well as their spelling seems to render them rather unsuitable for testing Polish learners of English. While checking the tests we noticed that all of the test-takers who attempted to spell the word “apartment” on the productive test resorted to the Polish spelling, which made it virtually impossible to judge whether or not they were familiar with the English equivalent at all. Post-test interviews revealed that only five out of 15 A2 learners knew the word “apartment”. When asked to translate the word “*mieszkanie*” into English, ten came up with the word “flat”, two came up with both “flat” and “apartment”, and three said they would normally use the word “apartment”. Only six students were familiar

working week”), which translates into Polish as “typowy” (“typical”). Post-test interviews confirmed that the learners were less familiar with the latter meaning of the word “normal”.

with the word “motor”. The remaining ones admitted to not knowing the word, but suspected it would be fairly similar to its Polish equivalent. As for the English equivalents of “*edukacja*” and “*wino*” (item difficulty coefficients of 0.8-0.93), the vast majority of subjects were able to provide correct answers during the interview.

Taking the above into consideration, we decided to replace two words, namely “motor” and “apartment”, with other nouns from relevant frequency levels. After careful consideration the word “motor” (1841st word in the BNC corpus, 5052 counts) was replaced by its synonym “engine” (1407th word in the BNC corpus, 6883 counts), and the word “apartment” (3612th word in the BNC corpus, 1901 counts) was replaced by a semantically related word “residence” (3779th word in the BNC corpus, 1789 counts) / “*przebywanie*”, “*miejsce zamieszkania*”, “*siedziba*” (counts quoted from frequency lists compiled by Kilgarriff on the basis of the BNC corpus, 2006).

Creation of the productive 1,000 Vocabulary Levels Test

The Laufer and Nation productive Vocabulary Levels Test (1999) does not contain a 1,000-word frequency test. For the purpose of the current study we developed and piloted the 1,000-word frequency test (its pilot results are presented in Tables 5.10 and 5.11 in the following section alongside the results of selected sections from Vocabulary Levels Test). Like other productive sections of Vocabulary Levels Test, the 1,000-word frequency test consists of nine nouns, six verbs and three adjectives (see Appendix F). Half of the test items were selected from the first 500 most frequent words, and the other half from the second 500 most frequent words.

Comparing the amended versions of the Vocabulary Levels Test

The noun-verb-adjective ratio in the receptive versions of the Vocabulary Levels Test is 5:3:2, which roughly corresponds to the 3:2:1 ratio of relevant parts of speech in the Thorndike and Lorge list (1944). Each receptive version of the Vocabulary Levels Test includes 15 nouns, nine verbs and six adjectives. Each productive version has a noun-verb-adjective ratio of 3:2:1, with nine nouns, six verbs and three adjectives at each frequency level. There is a certain overlap between receptive and productive Versions 1 and 2 of the 2,000 and 3,000-word frequency tests in terms of the target words tested. Altogether 130 words are tested in Versions 1 and 2, of which 62 are

tested both productively and receptively. The number of words tested receptively only and productively only is 58 and 10 respectively (see Table 5.8 below for a few examples and Tables A.4 to A.6 in Appendix F for the full lists).

Table 5.8. Examples of words included in the 2,000 and 3,000 Vocabulary Levels Test and their mode of testing

Mode of testing	Examples of target words	Test and item number
receptive	education (n) melt (v) original (adj) angel (n) dwell (v) odd (adj)	2,000-word level, Receptive 1, item 7 2,000-word level, Receptive 2, item 24 2,000-word level, Receptive 1, item 25 3,000-word level, Receptive 2, item 11 3,000-word level, Receptive 2, item 16 3,000-word level, Receptive 1, item 28
productive	deed (n) nurse (n) roots (n) acid (adj) career (n) snap (v)	2,000-word level, Productive 2, item 1 2,000-word level, Productive 2, item 3 2,000-word level, Productive 2, item 2 3,000-word level, Productive 1, item 2 3,000-word level, Productive 1, item 1 3,000-word level, Productive 2, item 14
both	charm (n) climb (v) slight (adj) lawn (n) scare (v) stable (adj)	2,000-word level, Receptive 1, item 11 2,000-word level, Productive 1, item 5 2,000-word level, Receptive 1, item 16 2,000-word level, Productive 2, item 10 2,000-word level, Receptive 2, item 29 2,000-word level, Productive 1, item 16 3,000-word level, Receptive 2, item 14 3,000-word level, Productive 1, item 8 3,000-word level, Receptive 1, item 16 3,000-word level, Productive 1, item 10 3,000-word level, Receptive 2, item 26 3,000-word level, Productive 1, item 17

Note: Receptive versions 1 and 2 were devised by Schmitt (2000) & Nation (2001) respectively. Receptive Version 2 was further modified by Schmitt et al. (2001). Productive versions 1 and 2 were devised by Laufer & Nation (1999).

Amended versions of the productive and receptive Vocabulary Levels Test were piloted on 15 A2 learners (male/female = 6/9; age range = 22-29, μ = 23.93, SD = 1.71) and 15 B1 learners (male/female = 5/10; age range = 23-26, μ = 23.87, SD = 0.99) studying English as a minor at the Higher School of Marketing in Silesia, Poland. The tests were administered in a pen and paper format, as we were not granted access to computer facilities in the college. Since a significant number of items were presented in identical sequences on the receptive and productive tests (see Table 5.9 below), the sets of items on the receptive test and the individual items on the productive test were rearranged randomly within each frequency band.

Table 5.9. Examples of identical sequencing of some items in the receptive and productive 2,000 and 3,000 Vocabulary Levels Test

Level	Target words	Item number			
		Receptive 1	Receptive 2	Productive 1	Productive 2
2,000 level tests	treasure	10		4	
	charm	11		5	
	lack	12		6	
	cream	13		7	
	wealth	14		8	
	pupil	15		9	
	climb	16			10
	examine	17			11
	surround	18			12
	connect	19			13
	wander	20			14
	limit	21			15
3,000 level tests	slender		25	16	
	stable		26	17	
	naked		27	18	
	normal		28		16
	supreme		29		17
	aware		30		18

The Vocabulary Levels Test data are normally distributed for both proficiency groups – the values of skewness and kurtosis are non-significant for each receptive and productive test (Tables 5.10 and 5.11 below). On average the A2 students obtained 75.9 % (29.6 points), 50% (15 points) and 30.6 % (9.2 points) on Version 1 of the receptive 1,000, 2,000 and 3,000 Levels Test respectively. The points obtained on Version 2 of the respective tests are 29.2, 14.47 and 9.13. The results of receptive Versions 1 and 2 are highly correlated at each of the three frequency levels – 0.963 for the 1,000-word level, 0.973 for the 2,000-word level and 0.836 for the 3,000-word level. Additionally, the differences between the means of the two versions are non-significant at all levels ($t_{1000} = 1.193, p > .05$; $t_{2000} = 1.586, p > .05$; $t_{3000} = 1.193, p > .05$).

The means of the productive 1,000, 2,000 and 3,000 tests are 13.40, 8.33/6.06 and 1.27/1.20 respectively. The productive 1,000 test has only one version, and to assess its reliability the test was re-administered after a two-week period. The data from the first administration is included in Table 5.10. The test-retest reliability for the 1,000 test is 0.951. The results of the productive 2,000 Version 1 and Version 2 are highly correlated (0.867, $p < .01$), but the difference between the two means is shown to be significant ($t = 3.90, p < .01$). The results on the 3,000 versions are too low to draw any firm conclusions about the extent to which these versions can be considered parallel. The results are bunched together and range between 0 and 2 points ($k = 18$). On the one

hand, the difference between the means is not significant ($t = 0.366$, $p > .05$), possibly due to the floor effect. On the other hand, the two sets of data are not correlated ($r = 0.05$), which indicates a lack of correspondence between scores obtained on both versions by the same individuals, again very likely due to the limited range of scores obtained on both versions.

Table 5.10. Piloting - descriptive statistics for parallel versions of the Vocabulary Levels Test (n=15, level=A2)

Test	No of items	Min	Max	Mean	SD	Skewness	Kurtosis
1,000 receptive 1	39	18	36	29.60	4.733	-0.723, ns	1.272, ns
1,000 receptive 2	39	18	35	29.20	4.246	-1.122, ns	2.408, ns
2,000 receptive 1	30	7	22	15.00	5.182	-0.124, ns	-1.684, ns
2,000 receptive 2	30	6	22	14.47	5.540	-0.228, ns	-1.545, ns
3,000 receptive 1	30	6	12	9.20	2.077	-0.086, ns	-1.608, ns
3,000 receptive 2	30	7	11	9.13	1.505	-0.259, ns	-1.495, ns
1,000 productive	18	10	17	13.40	2.229	-0.137, ns	-1.055, ns
2,000 productive 1	18	1	11	8.33	4.353	-0.165, ns	-0.708, ns
2,000 productive 2	18	2	10	6.06	3.173	-0.286, ns	-1.740, ns
3,000 productive 1	18	0	2	1.27	0.594	-0.091, ns	-0.171, ns
3,000 productive 2	18	1	2	1.20	0.414	1.672, ns	0.897, ns

Considering the extensive piloting of the receptive tests (Schmitt et al., 2001) and the results we obtained while testing the parallel versions of the test on our A2 students, we are fairly confident that the two versions are similar enough to measure vocabulary gains over time for learners of relatively low proficiency. As for the productive tests, the means of the 2,000 Version 1 and Version 2 have been shown to be statistically different when calculated on the basis on the A2 students’ responses. We fear, therefore, that using these productive tests for test-retest purposes could fail to show true gains in vocabulary knowledge among students of such low proficiency. As a result, we opted to use just one of the two versions. Version 2 was selected as it has fewer prompts that provide most of the word stem (see prompts in bold in relevant productive tests in Appendix F). With regard to test difficulty, bearing in mind the students’ poor performance on the 3,000 test (especially on the productive format), we felt that our initial decision to focus on levels 1,000 and 2,000 was justified.

The B1 students obtained 93.68 % (36.53 points), 84 % (25.20 points) and 73.78 % (22.13 points) on Version 1 of the receptive 1,000, 2,000 and 3,000 tests (see Table 5.11 below). The results of Version 2 appear very similar, namely 36.13, 25.40

and 22.40 points, which is confirmed by the fact that the differences between the means of the parallel tests at each frequency level are statistically non-significant ($t_{1000} = 1.146, p > 0.05$; $t_{2000} = -0.526, p > 0.05$; $t_{3000} = -0.694, p > 0.5$). Additionally, the two sets of results at each frequency level are highly correlated (0.766, 0.799 and 0.911 for the 1,000, 2,000 and 3,000 tests, all $ps < 0.01$). With regard to the productive tests, the B1 learners obtained 16.8, 12.33/10.87 and 9.20/8.87 on the 1,000, 2,000 and 3,000 Vocabulary Levels Test. The difference ($d=0.33$) between the means of the productive 2,000 tests is statistically non-significant ($t=1.824, p > 0.05$), as is the difference ($d=1.47$) between the means of the 3,000 tests ($t=0.357, p > 0.05$). In the former case, however, the value of significance is relatively close to the 0.05 cut-off point ($p=0.089$). The parallel tests' data, however, are not correlated (0.094 and 0.036). This finding is unexpected, considering the piloting carried out by Laufer and Nation (1999). It illustrates what a challenging task it is to create parallel versions of a test, and that the validation data from a group of learners of certain nationalities may not necessarily be applicable to learners with different linguistic backgrounds.

Table 5.11. Piloting - descriptive statistics for parallel versions of the Vocabulary Levels Test (n=15, level=B1)

Test	No of items	Min	Max	Mean	SD	Skewness	Kurtosis
1,000 receptive 1	39	34	39	36.53	1.685	0.140, ns	-0.909, ns
1,000 receptive 2	39	33	39	36.13	2.099	0.065, ns	-1.239, ns
2,000 receptive 1	30	20	28	25.20	2.242	-0.595, ns	0.451, ns
2,000 receptive 2	30	21	29	25.40	2.384	-0.185, ns	-0.567, ns
3,000 receptive 1	30	17	28	22.13	3.603	-0.209, ns	-1.086, ns
3,000 receptive 2	30	18	28	22.40	3.158	0.391, ns	-0.922, ns
1,000 productive	18	15	18	16.8	1.014	-0.493, ns	-0.598, ns
2,000 productive 1	18	7	17	12.33	2.870	-0.449, ns	-0.648, ns
2,000 productive 2	18	8	13	10.87	1.505	-0.320, ns	-0.742, ns
3,000 productive 1	18	5	13	9.20	2.624	-0.059, ns	-0.598, ns
3,000 productive 2	18	4	12	8.87	2.587	-0.487, ns	-0.844, ns

In the light of the above findings from a group of B2 students, we decided to use both versions of the receptive test, but only one version of the productive test, i.e. Version 2. Since in the main experiment the first and second administration of the receptive test were scheduled six months apart, we did not expect the learners would remember much from the first test administration when re-tested six months later. As for the test difficulty, the B1 learners appear to be familiar with the majority of the 1,000 most frequent word families (93.7 on the receptive versions and 93.3 on the

productive test). Further gains within this frequency band are likely to be very slight, if any at all, as many learners reach a plateau at about 95 % in the most frequent bands (Milton, 2009: 31). As a consequence, we feel the 2,000 and 3,000 levels are most suitable for B1 learners in our main experiment.

During the piloting of the receptive 1,000 Vocabulary Levels Test attention was also paid to those items on the test that draw on subjects' general knowledge, as in the two examples below:

Each society has the same rules.
You can go by road from London to New York.

The difficulty with such sentences arises out of the fact that without knowing the location of London and New York it is impossible for the test-taker to tell whether “you *can* go by road from London to New York” or *cannot*, despite being able to understand each word in that sentence. The general knowledge factor could, therefore, pose some problems for those test-takers who happen to be less familiar with some of the places and concepts incorporated in the test.

We identified seven sentences on the test that in our opinion could be particularly influenced by the general knowledge factor (see also Nation, 1993: 195). Four of the sentences feature in Version 1, while the remaining three are part of Version 2 (see Table 5.12). Items 3, 19 (Version 1) and 28 (Version 2) appear fairly easy for both proficiency groups, while items 9 and 24 (Version 1), 24 and 33 (Version 2) turn out to be rather difficult, but only for the A2 students. When calculated for the A2 test-takers, the item difficulty coefficients of the latter group of sentences range between 0.4 and 0.53. When calculated for the B1 students, the coefficients range between 0.93 and 1.0. The fact that these three sentences appear quite easy for the B1 test-takers indicates that adult Polish test-takers have the general knowledge required to provide the correct answers to these items, and that the proficiency level rather than general knowledge is a factor here. Post-test interviews seem to confirm this conclusion. The subjects who failed to provide correct answers did not understand some of the following words, i.e. “remain”, “still”, “society”, “made of”, “impossible” and “without”. One of the A2 subjects admitted to not being able to understand the construction “it is easy for someone to (verb)” despite being able to understand all the words involved. Post-test interviews reveal that when presented with Polish translations

of the seven sentences, the subjects did not have any difficulty providing correct answers.

Table 5.12. Piloting - item difficulty coefficients of selected items from the receptive 1,000 Vocabulary Levels Test

Test and item number	Test sentences	Item difficulty coefficients	
		A2 students	B1 students
Version 1, item 3	Some children call their mother Mama.	0.80	0.87
Version 1, item 9	It is easy for children to remain still.	0.53	0.93
Version 1, item 19	You can go by road from London to New York.	0.73	0.93
Version 1, item 24	A society is made of people living together.	0.53	0.93
Version 2, item 24	Each society has the same rules.	0.40	1.00
Version 2, item 28	Rome (Rzym) is an ancient city.	0.67	0.87
Version 2, item 33	It is impossible to live for a long time without water.	0.53	0.93

On a final note, even though the productive Vocabulary Levels Test is believed to be less susceptible to guesswork than its receptive equivalent (e.g. Webb, 2008: 81), it is not impervious to guessing. The characteristics of the prompts provided to elicit the target words on the productive test differ, which may have a bearing on the amount of guessing taking place. In Version 1 and 2 of the productive 2,000 and 3,000-word level tests the number of letters per prompt ranges from one to five. It is not unlikely that such a variable number of letters confronts the test-takers with varying degrees of difficulty, especially in those instances where the prompts provide most of the word stem, as they do in the following examples:

People were **whir**_____ around on the dance floor. (*whirling*)
This sweater is too tight. It needs to be **stret**_____. (*stretched*)

As mentioned above, only one version of the productive Vocabulary Levels Test was selected for test-retest purposes, as the results obtained on the two versions by the same individuals were not statistically equivalent. Version 2 was favoured over Version 1, as the former contained fewer prompts that provided most of the word stem (see Appendix F).

5.3.4. L1 Vocabulary Test

The design of the L1 Vocabulary Test is based on the productive Vocabulary Levels Test (fill-in-the-blanks with prompts). The test probes the productive knowledge of 17 verbs, 28 nouns and 9 adjectives (see Table A.10 in Appendix H for an

alphabetical list with English translations). A receptive component is not included, as it could lower the level of variability in the overall L1 test scores. More information about the selection and types of test items is presented in Chapter 6, Section 6.3.7.

Normality and distribution of scores

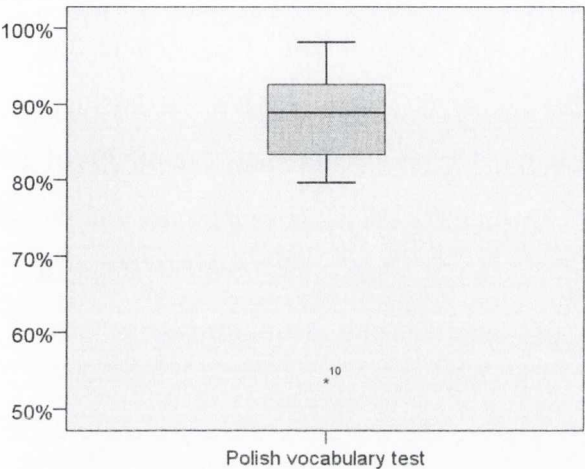
The L1 Vocabulary Test was first administered to three educated native speakers of Polish resident in Dublin, who were qualified to teach Polish as a foreign language. They reported having communicative knowledge of English, German and French. Two of them had arrived in Ireland within the previous 6 months, while the remaining one had been in Ireland for a year. Subject 1 and 3 completed all the items correctly, while Subject 2 left one blank empty. The item that posed a problem for Subject 2 was the word “pejoratywny” meaning “pejorative”⁵³ (sentence 39, Appendix H). As a result, a decision was made to add one additional letter to the prompt for that word (i.e. “j” in “pej...”). The L1 Vocabulary Test was subsequently piloted on 14 Polish nationals from Dublin (male/female = 7 / 7; age range = 23 / 35, $\mu = 29.5$, $SD = 3.976$; length of residence = 15.5 months, $SD = 8.08$; self-assessed level of English proficiency – A2 or B1). Prior to the test administration the participants were asked to indicate whether or not they had ever had dyslexia or other learning difficulties. None of the subjects admitted to having (had) learning difficulties. A computerised version of the test was prepared. To control for the possible effect of item order, the sentences were presented in random order. The order of sentence presentation was recorded for each subject.

The mean result on the test was 86.4% (min / max = 53.7 / 98.15, $SD = 10.665$). The minimum of 53.7 was surprisingly low, and we suspected that this might have been an outlier. As boxplots are very useful tools for pinpointing outliers, a boxplot displaying the distribution of the data was generated in SPSS 16.0. The generated boxplot confirmed our suspicion that that Case 10 (53.7%) was indeed an outlier (see Figure 5.1 below). In an attempt to get an insight into why Subject 10 obtained such a low score on the test, her test answers were re-ordered according to the sequence of presentation during the test. This revealed that Subject 10 filled in the first 31 missing

⁵³ A few test items – such as e.g. “alkowa”/“alcove”, “demistyfikacja”/ “demystification” “mozaika”/“mosaic” or “pejoratywny”/ “pejorative” – are cognates in Polish and English. We did not consider that to be a problem, though, as our subjects were unlikely to have the productive knowledge of the English equivalents due to the subjects’ relatively low level of English proficiency (A2 and B1).

words and left the remaining ones blank. Case 10 was subsequently excluded from further analysis.

Figure 5.1. Piloting – boxplot of the distribution of L1 Vocabulary Test scores (n=14)



After excluding the outlier (n=13), the L1 Vocabulary Test scores range from 79.63 to 98.12% (Table 5.13). The mean has gone up from 86.4 to 88.89 and now equals the median of the distribution. There is also a substantial reduction in standard deviation (from 10.665 to 5.238). A boxplot and a normal probability plot (see Appendix J) for the reduced set of data indicate no deviations from normality. The Kolmogorov-Smirnov coefficient and values of skewness and kurtosis are non-significant, which confirms the assumption of normality.

Table 5.13. Piloting – descriptive statistics for L1 Vocabulary Test scores (n=13)

L1 Vocabulary Test	mean (μ)	88.89 %
	min /max	79.63 / 98.12 %
	SD	5.238 %
	Kolmogorov-Smirnov test	0.2, <i>ns</i>
	Skewness	0.00, <i>ns</i>
	Kurtosis	-0.463, <i>ns</i>
	Guttman split-half reliability	0.91

The Guttman split-half reliability estimate has increased from .846 to .91 and now represents very high internal consistency reliability. In order to calculate a split-half reliability estimate, as previously, the scores were split into two halves. The split was a combination of a rational and random approach. First, test scores were arranged according to the part of speech they were testing – nouns, verbs and adjectives. Then

within each of these three groups scores were randomly assigned to either the 1st or 2nd half of the total score.

Item difficulty index

The average item difficulty when piloted on the group of 13 subjects is .863. The test was designed for native speakers of Polish, and hence most of the test items were expected to have an item difficulty coefficient between 0.9 and 1.0. 40.7% of test items have an item difficulty estimate of 1.0, while 48.2% have an estimate that ranges between 0.7 and 0.9. Crucially, there are a number of items on the test (11.1%) that turned out to be relatively difficult for the test-takers during the piloting stage (see Table 5.14 below).

Table 5.14. Piloting – L1 Vocabulary Test items with an item difficulty coefficient below 0.7 (n=13)

Polish word	Part of speech	Translation	Item difficulty
demistyfikacja	noun	demystification	0.385
pejoratywny	adjective	pejorative	0.385
rozłąka	noun	separation (e.g. of families)	0.538
oponować	verb	oppose	0.615
tarapaty	noun	plight, trouble	0.615
uwierzytelnić	verb	authenticate	0.692

The pilot study also revealed that despite the fact that the task instructions were given in Polish and that the subjects were informed explicitly that they could fill in each blank with only one word, on a few occasions more than one word was provided. In sentence 21, for example, where the word “oponował”/ “oppose” (3rd person singular, past tense, imperfective) was required, two subjects provided a reflexive verb “opowiedział się” / “speak (against)”. To prevent this problem from occurring in the main study one more letter was added to the prompt (i.e. “opon...”).

Chapter 6: Sample and Research Instruments

6.1. Introduction

The present chapter consists of two sections dealing with the sample characteristics and the research instruments. The first section (6.2) presents the characteristics of the adult Polish learners of English who participated in the current study as well as some background information on Polish migration to Ireland prior to the year 2008. It also provides some information on the English language course that was organised in Trinity College Dublin in order to gather the data for the study. The second section (6.3) gives a comprehensive description of the final version of our research instruments.

6.2. Sample and language course description

6.2.1. Polish migration to Ireland prior to 2008

Our data collection started in January 2008, approximately three and a half years after the May 2004 enlargement of the European Union (EU). At the time of accession the Polish labour market was performing rather poorly. The unemployment rate in Poland, nearly 20%, was the highest unemployment figure of all A8 countries (Drinkwater, Eade & Garapich, 2006: 4). Wages were very much on the low side, with average hourly labour costing approximately 4.41 Euro in Poland, as compared to e.g. 27.09 Euro in West Germany (Sinn, 2004: 2). High unemployment rates in accession states and considerable wage differences between EU-15 countries and new EU members created substantial pressures for migration upon accession.

The fact that Ireland and the UK offered open labour market access to citizens of Poland and other accession states has been a factor contributing to the largest ever waves of migration in those areas (e.g. Salt & Rees, 2006). The immigration restrictions imposed by the majority of EU-15 countries resulted in redirecting the flow of migrant traffic to those EU-15 states that had more liberal immigration policies (Barrell, FitzGerald & Riley, 2007; Boeri & Brücker, 2005). Ireland and the UK were among open labour market EU-15 members that received larger EU8 inflows than

expected. Having analysed the Irish Quarterly National Household Survey and other available sources, Barrel et al. (2007: 4) report that in the months following the EU enlargement out of 313,800 Polish emigrants 37,900 moved to Ireland and 167,500 moved to the UK.

According to the 2006 Census Polish nationals were the dominant migrant group among migrants from the new EU states, followed by Lithuanians (24,638) and Latvians (13,319). The exact number of Polish migrants in the years after the EU accession is not known, but the figures emerging from the 2006 Census indicate that there were at least 63,276 Poles living in Ireland at the time when the poll was conducted. However, informed estimations from a number of sources put the figure much higher, closer to 200,000 (see e.g. Holmquist 2006: 1). This suddenly growing number of Polish migrants in Ireland opened up new opportunities for research on adult learning of English as a second language in an English-speaking country.

6.2.2. Sample description

The sample for the main study comprises 60 adult Polish learners of English resident in Ireland, 30 of which are at the elementary (A2) level and 30 at the pre-intermediate (B1) level. The proficiency levels were defined in accordance with Common European Framework of Reference (Council of Europe, 2001) and the targeted age range for our research project was 25-35. The native speakers of Polish who expressed willingness to participate in our project were offered a six-month English language course at Trinity College Dublin free of charge. We anticipated a certain amount of dropout from the course due to motivational factors or other unforeseen circumstances such as e.g. the need to re-locate in search of employment. To ensure we would obtain longitudinal data from at least 30 students at each proficiency level, we recruited as many as 104 subjects from the subject panel of the Polish Diaspora Project (see Chapter 1). As predicted, over 40 % of the learners who signed up for the course had to be excluded from our analyses – a few individuals reported having hearing problems or dyslexia, some subjects dropped out before completing the course, while others failed to attend one of the testing sessions.

The longitudinal data we obtained in the course of this study came from 24 males and 36 females, with 13 males at the A2 level and 11 at the B1 level (see Table

6.1 below). The mean age for the A2 and B1 groups was 29.3 and 30.4 respectively. 65 % of our subjects had studied at tertiary level, 6.7 % had vocational education and 28.3 % had completed secondary education. The mean length of residence in Ireland was 17 and 24 months for A2 and B1 students respectively. 17% of A2 students and 27% of B1 students had their first encounter with English at a tertiary level. 54% of A2 students and 73% of B1 had some contact with English either in a primary or secondary school. 29% of A2 students started learning English after their arrival in Ireland. The average length of learning English for A2 and B1 students prior to participating in the study was 2 and 4 years respectively. None of our subjects attended any other English language course for the duration of this study.

Table 6.1. Background information about participants

Group	A2 (n=30)				B1 (n=30)			
Gender	13 males 17 females				11 males 19 females			
Age (years)	M = 29.3 SD = 4.091				M = 30.4 SD = 3.654			
Residence in Ireland (months)	M = 17.16 SD = 7.61				M = 24 SD = 11.76			
Context of first exposure to English	primary	8 ^A	26.67 %		primary	8	26.67 %	
	secondary	8	26.67 %		secondary	7	23.33 %	
	vocational	-	0 %		vocational	-	0 %	
	tertiary	5	16.67 %		tertiary	8	26.67 %	
	other (e.g. private tuition)	9	30%		other (e.g. private tuition)	7	23.33 %	
Education	secondary	14	46.67 %		secondary	3	10 %	
	vocational	3	10.0 %		vocational	1	3.33 %	
	tertiary	13	43.33 %		tertiary	26	86.67 %	
Other foreign languages	Russian	15	50 %		Russian	13	43.33 %	
	German	14	46.67 %		German	20	66.67 %	
	French	4	13.33 %		French	4	13.33 %	
	Italian	1	3.33%		Italian	2	6.67 %	
	Spanish	1	3.33%		Dutch	1	3.33 %	
Number of other foreign languages per student	one	25	83.33 %		one	23	76.67 %	
	two	5	16.67 %		two	7	23.33 %	

^ANumber of students as well as percentages are specified.

All of the students who completed the course were planning to stay in Ireland for at least one more year after the course termination, and were highly motivated to keep improving their knowledge of English. Over 40 % of the participants expressed willingness to stay in Ireland permanently, and only 5 % said they were considering their stay in Ireland as a short-term arrangement. The reasons for learning English were

numerous – mainly to communicate with English speakers and make new friends, but also to get hired for a better job, and in a few cases to feel part of Irish society.

6.2.3. English language course description

During an introductory meeting they were familiarized with the timeline and objectives of the course. They were also presented with requirements for obtaining a credit from the course – types of homework assignments and tests to be completed, number of absences allowed, etc.

In the English language course due attention was paid to both paradigmatic and syntagmatic aspects of word knowledge (Lewis, 2002). The textbooks used in the A2 group included *New Headway: Elementary* (Soars & Soars, 2006), *English Vocabulary Organizer: 100 topics for self-study* (Gough, 2000), and *Essential Grammar in Use: Elementary* (Murphy, 2007), and those used in the B1 group included *Total English: Intermediate* (Clare & Wilson, 2006), *Enterprise 3: Grammar* (Evans & Dooley, 1999) and *English Collocations in Use* (McCarthy & O'Dell, 2005). Additional teacher materials were also incorporated. The students were divided into four groups, and were taught by the author of the current thesis⁵⁴, a teacher of English as a Foreign Language with substantial teaching experience. The students had two 45-min lessons a week blocked together.

6.3. Instruments

6.3.1. Background questionnaire

Our background questionnaire (see Appendix B) was partially modelled on the questionnaires used by Fouser (2000) and Bielska (2002) and that used in the Modern Languages Research Project at Trinity College Dublin (Singleton, 1999). It is comprised of two main sections, one eliciting general information about the subjects and the other information on their English language learning experience and patterns of

⁵⁴ My decision to teach the TCD course was motivated by a number of factors. Lack of groups of Polish learners who would be suitable in terms of English proficiency and age (and whose English language teacher would be willing to co-operate) contributed to the decision to organise and teach a language course myself. Also, I felt it was essential to ensure that balance would be maintained between teaching vocabulary and collocations, on the one hand, and teaching other language skills and subskills, on the other. It was also very important to ensure that the words and collocations selected for an end-of-course test would be introduced and recycled during the course.

exposure and use. Apart from routine questions about the subjects' age, sex or educational background, it addresses the following aspects:

- a) self-assessment of general English proficiency – a six-point scale ranging from “A1 – elementary” to “C2 – advanced”,
- b) self-assessment of specific skills and subskills in English – a five-point scale ranging from “1 – excellent/near nativelike” to “5 – poor”,
- c) areas of ease and difficulty in relation to learning English language skills and subskills – a five-point scale ranging from “1 – very easy” to “5 – very difficult”,
- d) goals in studying English – a multiple-choice format.

At an introductory meeting held prior to the commencement of the English language course our subjects were requested to fill out a computerised version of the questionnaire along with a consent form (Appendix A). They were informed that the data obtained during the current study would be treated as research findings for presentations and publications, and that their names and other personal details would not be disclosed in any publications, discussions or correspondence related to this study. They were also informed they could withdraw from the research project at any stage without any ramifications. The participants were made aware of the fact that the English course was part of a bigger project sponsored by Irish Research Council for Humanities and Social Sciences on second language acquisition and native language maintenance in the Polish diaspora in Ireland and France.

6.3.2. Serial Nonword Recall

The Serial Nonword Recall task consists of three sets of nonword lists of varying lengths (see Appendix C). Each set of lists was constructed from a pool of 20 CV⁵⁵-CCV-CV nonwords. All of the nonwords are pronounceable phonological sequences according to Polish phonotactics. The nonword lists were recorded in the TCD phonetics laboratory and subsequently edited in Audacity 1.2.6⁵⁶ in order to ensure that each phoneme's pronunciation was clear and that pauses between nonwords were identical.

⁵⁵ “C” stands for a consonant and “V” for a vowel.

⁵⁶ Audacity 1.2.6 can be retrieved from <http://audacity.sourceforge.net>.

As mentioned in Chapter 5 (Section 5.3.1), the syllables used to create the experimental nonwords were sourced from a Polish language dictionary (Polish Scientific Publishers PWN – CD-ROM edition) and a frequency dictionary of Polish (Kurcz et al., 1990), and are as follows:

- 20 initial CV (consonant-vowel) syllables: bi, by, ce, do, fe, fo, ga, gi, je, ju, ko, la, le, mu, ne, ra, ro, se, wo, wy,
- 20 medial CCV syllables: ble, bru, cmo, dry, flo, gli, gro, klo, kra, kse, pli, pro, pty, ska, smu, sta, tru, wle, zbi, zna,
- and 20 final CV syllables: by, ca, do, dy, go, ka, ki, ko, la, le, mi, ne, no, pu, ra, sa, sy, ta, ty, wy.

The positions (initial, medial or final) in which the above-listed syllables appear in three-syllable Polish words are reflected in the constructed nonwords. Each constituent syllable is used only once within this pool of nonwords, and there are no consonant or vowel repetitions within a nonword.

Wordlikeness ratings, reflecting the degree to which 20 native speakers of Polish view the constructed nonwords as resembling real Polish words, range from 1.3 to 2.35 (see Table 6.2). For more information about the process of nonword construction and selection please consult Chapter 5.

Table 6.2. Wordlikeness ratings for nonwords in the Serial Nonword Recall task

Nonwords	Wordlikeness ratings (means)
	1 – very unlikely to pass for a real Polish word 5 – very likely to pass for a Polish word
biznale	2.2
bywlera	2
ceptyko	1.7
dokrapu	1.85
fezbino	2.1
foksela	1.9
gaflosy	2
gibruta	2
jestaby	1.7
jukloty	2.2
kosmuca	1.7
lagroki	1.85
letrumi	1.65
mudrygo	1.6
neglika	2.1
raplido	1.3
roblewy	2
seprody	1.95
woskane	2.35
wycmosa	1.8

Nonwords	Wordlikeness ratings (means)
	1 – very unlikely to pass for a real Polish word 5 – very likely to pass for a Polish word
Descriptive statistics	min / max = 1.3 / 2.35 $\mu = 1.89$ SD = 0.25

Each of the three sets of nonword lists is made up of 15 lists of the same length. Set one, two and three contain 2-item, 3-item and 4-item lists respectively (Appendix C). The nonwords assigned to each list start with different consonants and they do not sound alike. The position of each nonword is controlled to ensure that they occur in a variety of positions within the pool of lists. The frequency of occurrence of nonwords within each set of lists ranged from 6 to 7($\mu = 6.75$, SD = 0.44). We made an effort to ensure that the distribution of nonwords in each set was comparable (Table 6.3 below).

Table 6.3. Frequency of nonword occurrence in 2-, 3- and 4-item lists in the serial recall task

Nonwords	Frequency in 2-item lists	Frequency in 3-item lists	Frequency in 4-item lists	Total frequency
biznale	2	2	3	7
bywlera	1	3	3	7
ceptyko	1	3	3	7
dokrapu	2	2	3	7
fezbino	2	2	3	7
foksela	2	2	3	7
gaflosy	2	2	3	7
gibruta	1	3	3	7
jestaby	2	2	3	7
jukloty	2	2	3	7
kosmuca	1	2	3	6
lagroki	1	2	3	6
letrumi	1	3	3	7
mudrygo	1	2	3	6
neglika	2	2	3	7
raplido	1	2	3	6
roblewy	2	2	3	7
seprody	2	2	3	7
woskane	1	3	3	7
wycmosa	1	2	3	6
Descriptive statistics	$\mu = 1.5$ SD = 0.51	$\mu = 2.25$ SD = 0.44	$\mu = 3$ SD = 0	$\mu = 6.75$ SD = 0.44

At the beginning of the English language course (January 2008) and after the termination of the course (July and August 2009) our subjects were tested on a one-to-one basis in a language laboratory in Trinity College Dublin. The participants were instructed to repeat each sequence of nonwords immediately after its presentation in

exactly the same order in which the items had been presented. No corrective feedback was provided during testing. Before the commencement of the task the subjects were involved in a 5-minute trial session during which they were familiarized with the task procedure.

The measure of performance on this test was the number of correctly repeated lists across all list lengths tested. A repeated list had to contain no mistakes to be accepted as correct. The testing was discontinued if a subject failed to repeat seven out of 15 lists of a certain length. The maximum possible score was 45.

6.3.3. Serial Nonword Recognition

Our serial recognition task consists of three sets of nonword lists of varying lengths. Each set of lists has been constructed from a pool of 20 Polish-sounding CV-CCV-CV nonwords, other than those used in the Serial Nonword Recall task (see Chapter 5, Section 5.3.1 for more information on the procedure of nonword selection). The syllables from which our experimental nonwords in the recall task are made up include the following:

20 initial consonant-vowel (CV) syllables: ba, bo, ce, da, fi, ga, gu, je, ju, ka, li, me, na, pe, ru, ry, so, to, we, zu,

20 medial CCV syllables: ble, bra, cmo, dli, dmua, dra, fro, gry, gwi, klu, kso, mro, pli, pry, ska, snu, spa, sta, zdo, zma,

and 20 final CV syllables: ba, by, ce, da, do, ga, gi, ka, ko, la, le, mi, na, ny, pu, ra, sy, to, ty, wi.

Wordlikeness ratings (Table 6.4), reflecting the degree to which the constructed nonwords are viewed as resembling real Polish words, range from 1.45 to 2.35 ($\mu = 2.08$, $SD = 0.24$). This indicates that the created nonwords resemble Polish words but are rather unlikely to pass for real words.

Table 6.4. Wordlikeness ratings for nonwords in the Serial Nonword Recognition task

Nonwords	Wordlikeness ratings (means)
	1 – very unlikely to pass for a real Polish word 5 – very likely to pass for a Polish word
bafroce	2.25
bostagi	1.95
ceplira	2.1
dagryko	2.1
fimrosy	2.35
gadlipu	2.05
gubrawi	2.3
jedmuda	2.3
jubleny	2.35
kacmoba	1.85
lizmato	1.9
meksona	2.25
nazdoty	1.9
pegwika	2.35
ruspale	1.7
rysnudo	2.25
sopryla	2.25
toskaby	2.05
wekluga	1.9
zudrami	1.45
Descriptive statistics	min / max = 1.45 / 2.35 $\mu = 2.08$ SD = 0.24

Our Serial Nonword Recognition task is comprised of 30 pairs of nonword lists with 10 pairs of nonword sequences at each of three list lengths, that is at 4-item, 5-item and 6-item lengths (see Appendix D). The nonwords assigned to each list do not sound alike. Just as in the recall task, nonwords occur in a variety of positions within the pool of lists. The frequency of occurrence of nonwords within each set of lists ranges from 7 to 8 ($\mu = 7.5$, $SD = 0.51$; see Table 6.5 below).

Table 6.5. Frequency of nonword use in 4-, 5- and 6-item lists in Serial Nonword Recognition

Nonwords	Frequency in 4-item lists	Frequency in 5-item lists	Frequency in 6-item lists	Total frequency
bafroce	2	2	3	7
bostagi	2	2	3	7
ceplira	2	2	3	7
dagryko	2	3	3	8
fimrosy	2	3	3	8
gadlipu	2	2	3	7
gubrawi	2	3	3	8
jedmuda	2	3	3	8
jubleny	2	2	3	7
kacmoba	2	2	3	7
lizmato	2	3	3	8
meksona	2	3	3	8
nazdoty	2	2	3	7
pegwika	2	3	3	8
ruspale	2	2	3	7
rysnudo	2	2	3	7
sopryla	2	3	3	8
toskaby	2	3	3	8
wekluga	2	2	3	7
zudrami	2	3	3	8
Descriptive statistics	$\mu = 2$ SD = 0	$\mu = 2.5$ SD = 0.51	$\mu = 3$ SD = 0	$\mu = 7.5$ SD = 0.51

The testing took place in the same language laboratory in Trinity College Dublin. The subjects were tested in January and re-tested in February 2008. Considering the high number of testing sessions that had to be conducted, this time five students were tested at a time. The subjects were instructed to listen to 30 pairs of nonword lists through high-quality headphones. 10 pairs of nonword sequences were prepared at each of three list lengths, namely at 4-item, 5-item and 6-item lengths. At each list length, five of the ten pairs of nonword sequences were identical and the remaining five shared exactly the same nonwords but two of the nonwords in question were transposed in the second sequence. The initial and final nonwords never changed their position. A special webpage was developed for collecting responses. The students were instructed to listen to each set of lists and tick either “Same”, “Different” or “Not sure” on the relevant webpage. The subjects were told they would listen to each set only once. The measure of the performance on this test was the number of correct responses across all list lengths tested. The maximum possible score was 30.

6.3.4. Gitsaki Collocation Tests

Gitsaki (1999) developed three collocation tests, each designed for a different English proficiency level. Two of these tests are employed in our main study to

investigate the development of controlled productive general knowledge of collocations in A2 and B2 students. The piloting process and minor alterations to the test are presented in Chapter 5, Section 5.3.2.

The collocation test for beginners (referred to as the A2 Gitsaki Collocation Test) consists of 50 test items. The test contains 39 items that can be classified according to the BBI typology (Benson et al., 1986; see Section 4.3.2) as 24 examples of grammatical collocations and 15 examples of lexical collocations. The remaining 11 test items are phrasal verbs and are not part of the BBI typology. The grammatical collocations include the following four types: one noun + preposition phrase (G1), eight preposition + noun phrases (G4), five adjective + preposition phrases (G5), ten verb + preposition phrases (G8-D). The lexical phrases include five types, that is six verb + noun phrases (L1), one adjective + noun phrase (L3), one noun + verb phrase (L4), six noun + (of) + noun phrases (L5) and one verb + adverb phrase (L7).

The collocation test for intermediate students (referred to as the B1 Gitsaki Collocation Test) consists of 65 test items. The test contains 52 items that can be classified as 36 examples of grammatical collocations and 16 examples of lexical collocations according to the BBI typology (Benson et al., 1986). The remaining 13 test items are phrasal verbs and do not feature the BBI typology. The grammatical collocations include the following four types: five noun + preposition phrase (G1), 17 preposition + noun phrases (G4), five adjective + preposition phrases (G5), nine verb + preposition phrases (G8-D). The lexical phrases include four types as well, namely ten verb + noun phrases (L1), two noun + (of) + noun phrases (L5), one adverb + adjective phrase (L6), and three verb + adverb phrases (L7).

The collocations tested are two-word units with the co-occurring item appearing within 3 words to the left or right of the node. All of the collocations are listed in at least one of the following dictionaries: “*Oxford Collocations: Dictionary for Students of English*” (2002), “*The BBI Combinatory Dictionary of English*” (Benson et al., 1986), “*The BBI Dictionary of English Word Combinations*” (Benson et al., 1997) or “*Oxford Advanced Learner’s Dictionary*” (Hornby, 2005). Additionally, all of them appear in the BNC corpus. The frequencies of grammatical collocations range from 26 to 16090 (including phrasal verbs). The frequencies of lexical collocations are in the

range between 7 and 5000 (see Table 6.6 below and Tables A.1 and A.2 in Appendix E for more detail).

Table 6.6. The BBI types and BNC frequencies of occurrence of English collocations in Gitsaki tests

Test	Group of collocations	Type of collocation	No of test items	BNC frequency min	BNC frequency max	Overall frequency range
A2	grammatical collocations	noun + prep (G1)	1	306	306	45 - 16090
		prep + noun (G4)	8	847	13945	
		adj + prep (G5)	5	170	6443	
		verb + prep (G8-D)	10	45	16090	
	phrasal verbs	phrasal verb	11	427	5000	427 - 5000
	lexical collocations	verb + noun (L1)	6	22	174	16 - 1417
		adj + noun (L3)	1	107	107	
		noun + verb (L4)	1	16	16	
		noun + (of) + noun (L5)	6	20	1417	
		verb + adverb (L7)		1415	1415	
B1	grammatical collocations	noun + prep (G1)	5	26	1041	26 - 16090
		prep + noun (G4)	17	85	13945	
		adj + prep (G5)	5	31	5535	
		verb + prep (G8-D)	9	26	16090	
	phrasal verb	phrasal verb	13	83	5000	83 - 5000
	lexical collocations	verb + noun (L1)	10	7	5000	7 - 5000
		noun + (of) + noun (L5)	2	266	308	
		adv + adj (L6)	1	3862	3862	
		verb + adverb (L7)	2	14	1415	

As the above figures indicate, the number of collocations in the tests that involve filling in a preposition is rather high. It reaches 35 on the A2 test and 49 on the B1 test. Gitsaki herself does not provide any explanation as to why both tests include such a high number of prepositions. Ideally, the number of prepositions and other parts of speech to be filled in should be more balanced. Since both tests contain very frequent as well as less frequent word combinations (e.g. “on one’s way” vs. “take shorthand”), we decided to refrain from making major alterations with regard to the proportions of collocation types. It should be noted at this stage that there may be a difference in performance according to whether the tested item is to the left or the right of the node. This issue, however, is beyond the scope of the present study.

The measure of performance on the collocation tests was the overall score obtained by each student, and the gain in collocational knowledge between Time 1 and Time 2. The test was scored on a dichotomous scale (0,1), and minor spelling and grammatical mistakes were disregarded. Due to the fact that the number of items on the A2 and B1 tests differed, the obtained scores were turned into percentages.

6.3.5. Vocabulary Levels Test

General information

The Vocabulary Levels Test, a measure of L2 vocabulary size, comprises separate sections to assess the respective sizes of English receptive and (controlled) productive vocabulary at different levels of word frequency. The test was developed by Paul Nation (1983; 1990; 2001b) and colleagues (Laufer & Nation, 1999; Schmitt, 2000; Schmitt et al., 2001), and was made available free of charge for classroom use. Meara (1996: 38) refers to it as probably the “nearest thing we have to a standardized test in (English) vocabulary” (also Schmitt, 2000: 174). Read (2000: 118) considers it a very useful tool in estimating vocabulary size, particularly in the L2 context. Instead of giving a single estimate of overall vocabulary size this test battery estimates vocabulary knowledge at a few frequency levels: 1,000, 2,000, 3,000, 5,000 and 10,000 as well as at the University word level.

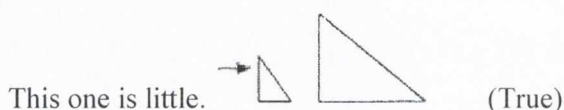
The first five levels were defined in relation to Thorndike and Lorge’s list (1944), the General Service List (West, 1953) and Kučera & Francis (1967). The words in the University word level were sampled from the University Word List (Xue & Nation, 1984). The original version of the University word level was constructed on the basis of a specialised list compiled by Campion and Elley (1971), as the Xue and Nation’s list was not available at the time. The word frequency levels addressed in our research project include the 1,000, 2,000 and 3,000 levels, and henceforth we shall focus on these tests only. For the purpose of our research a computerised version of the Levels Test was developed⁵⁷. Any departures from the original layout by Nation and colleagues will be brought to the reader’s attention.

There are a number of versions of the receptive and productive Levels Test. As described in Chapter 5 (Section 5.3.3), our project makes use of Version 1 and 2 of the receptive test (Schmitt et al., 2001) and Version 2 of the productive test (Laufer & Nation, 1999).

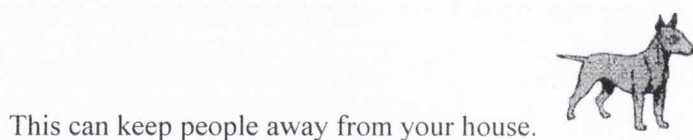
⁵⁷ Services provided by SurveyMonkey (<http://www.surveymonkey.com>) were used.

1,000-word frequency level tests

The 1,000-word frequency level test is a measure of receptive vocabulary knowledge and it is available in two parallel formats (Nation, 1993 & 2001). Each version consists of 39 short statements to be classified as either true or false, as the following example illustrates:



When designing the test, an attempt was made to ensure that the context words were of higher frequency than the words tested, e.g. “little” is less frequent than “this”, “one” or “is”. Whenever it was not possible to avoid a lower frequency word, a picture was used, as in the following sentence:



The grammatical complexity of the sentences was kept to a minimum. In most cases the sentences involve one clause only. However, in a few instances two-clause sentences appear in the test as well. Nation (1993) reports that the use of these sentences was unavoidable and that other two-clause items, those with low difficulty coefficients, were removed after the test was piloted.

In the period of time preceding the current study, there was no productive Levels Test for the 1,000-word frequency band available. Therefore, we developed the productive component and modelled it on the existing productive test by Laufer and Nation (1999). The test development and piloting is presented in Chapter 5, Section 5.3.3.

2,000 and 3,000-word frequency levels tests

The 2,000 and 3,000-word frequency levels have both receptive (Nation, 2001b modified by Schmitt et al., 2001; Schmitt, 2000) and productive components (Laufer & Nation, 1999) and are available in two parallel versions. In the receptive component test-takers are required to match words with definitions, where the definitions are the test items rather than the words themselves (the word-definition ratio at each level is 60 to 30). The ratio between nouns, verbs and adjectives in the stratified sample from the Thorndike and Lorge list is 3:2:1 (Schmitt et al., 2001: 58), which is roughly maintained in the Levels Test. Each frequency level contains 5 noun clusters, 3 verb clusters and 2 adjective clusters. The format of each cluster is presented below:

You must choose the right word to go with each meaning. Write the number of that word next to its meaning.

birth	_____ game
dust	_____ winning
operation	_____ being born
row	
sport	
victory	

The definitions within clusters are kept short to minimize the reading element and to allow for more vocabulary items to be tested within one testing session. To ensure that the test-takers' ability to demonstrate their knowledge of the target items is not compromised by their failure to understand some of the defining words, in most instances the words used in definitions are more frequent from the target vocabulary items. The defining words are essentially drawn from the 1,000 and 2,000 levels (for more detail see Nation, 1990: 264).

Each cluster of the receptive 2,000 and 3,000 Levels Test is designed with a view to eliminating aids to guessing. First, the target words appear in a decontextualized format, which prevents the making of inferences from the context about their meaning. Second, the words within each cluster are presented in alphabetical order, while the definitions appear in order of length. Third, the selection of the target vocabulary items from respective frequency bands is random. Also, the task instructions encourage the test-takers to leave empty blanks whenever they are not sure about the correct answer rather than guessing. In order to prevent learners from

leaving blanks empty by mistake, our computerised version of the Levels Test requires choosing one of the multiple-choice options alongside each definition before moving to the following cluster. The multiple-choice answers include A to F (denoting the target words) as well as the Not Sure option (unlike the original test). Our computerised version used the following item format:

- A. birth
- B. dust
- C. operation
- D. row
- E. sport
- F. victory

- 1. game
- 2. winning
- 3. being born

A
B
C
D
E
F
Not Sure

In the productive component of the Levels Test (Laufer & Nation, 1999) the test takers are presented with 18 sentences at each frequency level and are required to insert a missing word. A variable number of initial letters is provided to ensure that there is only one correct word that fits the gap, as in the following examples from the 3,000 word level:

The **pro**_____ of failing the test scared him.
Before writing the final version, the student wrote several **dra**_____.

The characteristics of the prompts provided and the parts of speech tested and are presented in Chapter 5, Section 5.3.3.

By Laufer and Nation’s own admission (ibid.: 37), all the test items were sourced from three parallel receptive versions developed by Schmitt rather than selected randomly from relevant frequency bands. Since not all versions of the productive and receptive Levels Test are statistically equivalent, only some of them are recommended for test-retest purposes (i.e. Version 1 and Version 2 used in this study). As a result, when parallel versions of the receptive and productive Levels Test are used

together, some items end up being tested both receptively and productively, while others are tested in one mode only (see Section 5.3.3 in Chapter 5, and Tables A.4, A.5 and A.6 in Appendix F).

The measure of performance on the Vocabulary Levels Test was the overall score obtained by each student, and the gain in vocabulary knowledge between Time 1 and Time 2. The test was scored on a dichotomous scale (0,1) and in the productive component grammatical and minor spelling mistakes were disregarded. Due to the fact that the number of items on the receptive and productive components differed, the obtained scores were turned into percentages.

6.3.6. Progress Vocabulary and Collocation Tests

The end-of-course test probes the knowledge of vocabulary and collocations taught and recycled during the course. Each item had to be recycled more than once to be included in the test. The distractors on the receptive tests were words that students encountered during the course.

The vocabulary component consists of two sections: a) 30 multiple-choice items that involved matching words with definitions (modelled on the receptive Levels Test format), and b) 18 fill-in-the-blank sentences (modelled on the productive Levels Test format; one-letter prompts provided). The receptive test includes 15 nouns, nine adjectives and six verbs. The productive test has the noun-verb-adjective ratio of 3:2:1, and it includes nine nouns, six verbs and three adjectives. The words tested in the productive test were selected from the pool of 30 words featuring on the receptive test (see Table 6.7 below and Table A.7, in Appendix G).

Table 6.7. Examples of words tested in the A2/B1 Progress Vocabulary Tests and their mode of testing

Mode of testing	Test	Word	Part of speech	Receptive test	Productive test
receptive and productive	A2	alone	adj	23	16
		begin	verb	15	10
		boss	noun	28	1
		chef	noun	9	2
	B1	affair	noun	27	1
		approve	verb	17	10
		archery	noun	5	2
		attempt	verb	15	11
receptive only	A2	accountant	noun	8	—
		boring	adj	20	
		drawing	noun	6	
		friendly	adj	19	
	B1	affordable	adj	24	—
		article	noun	26	
		charm	noun	7	
		columnist	noun	30	

The collocation test comprises three sections: a) 30 decontextualized multiple-choice questions (each containing one correct collocation and one pseudo-collocation), b) 30 collocation accuracy judgement sentences with a highlighted collocation each (the students were instructed to tick either “C/correct” or “I/incorrect”, and if “I/incorrect” the correct collocation was expected to be provided), and c) 30 fill-in-the blank sentences (with no prompts).

The A2 test includes six types of grammatical collocations and five types of lexical collocations (see Table 6.8 below and Table X in Appendix). The grammatical collocations comprise 23 prep + noun phrases (G4), three adj + prep phrases (G5), one verb phrase allowing dative movement (G8-A), one verb + to-inf phrase (G8-E), four verb + prep phrases (G8-D), and one verb + gerund phrase (G8-J). The lexical collocations include 27 verb + noun phrases (L1), 13 adj + noun phrases (L3), 15 noun + (of) + noun phrases (L5), one adverb of degree + adj phrase (L6), and one verb + adverb phrase (L7). The number of items within each category reflects the content of the course.

Table 6.8. The BBI types of collocations tested in the A2/B1 Progress Collocation Tests

Test	Group of collocations	Type of collocation	Example	Item no		
				Task 1	Task 2	Task 3
A2	Grammatical collocations	prep + noun (G4) adj + prep (G5) verb - allowing dative movement (G8-A) verb + to-inf (G8-E) verb + prep (G8-D) verb + gerund (G8-J)	at lunchtime angry at write a letter to sb learn to drive listen to stop raining	25 15 14	 10 23	5
	Lexical collocations	verb + noun (L1) adj + noun (L3) noun + (of) + noun (L5) adverb of degree + adj (L6) verb + adverb (L7)	cross the road safe driver bar of soap pretty easy work part-time	9 26 1	 26 28	
B1	Grammatical collocations	noun + prep (G1) prep + noun (G4) adj + prep (G5) verb - allowing dative movement (G8-A) verb + prep (G8-D) verb + gerund (G8-G) verb + object + inf (G8-I)	view on (the garden) in contact crazy about lend sb a book belong to avoid taking let sb go	 28 15 17	23	11 22 12
	Lexical collocations	verb + noun (L1) adj + noun (L3) noun + (of) + noun (L5) adverb of degree + adj (L6)	attend a lecture cramped house head of the household bang up-to-date	29	6 29 1	

The maximum score was 48 and 90 points in the vocabulary and collocation sections respectively. The test was piloted on three English native speakers resident in Ireland and minor adjustments were introduced.

6.3.7. L1 Vocabulary Test

To the best of our knowledge there is no standardized L1 Vocabulary Test available that could differentiate between native speakers of Polish in terms of their L1 proficiency. The design of our test is based on the productive Vocabulary Levels Test (fill-in-the-blanks with prompts). The test probes the productive knowledge of 17 verbs, 28 nouns and 9 adjectives (see Table A.10 in Appendix H for an alphabetical list with English translations). A receptive component is not included, as it could lower the level of variability in the overall L1 test scores. In order to increase the test’s power to detect linguistic differences between native speakers of Polish, the number of items on the test was increased from 18 to 54 (three times as much as in the Vocabulary Levels

format). The test items and the contexts in which they appeared in the L1 Vocabulary Test were selected from the online PWM Polish corpus⁵⁸.

The target Polish words have a frequency of less than 600 in the IPI PAN corpus, 2nd edition (Dębowski et al., 2006). The IPI PAN corpus is a 250 million word collection of samples of written Polish language, and it comprises extracts from newspapers, academic books, popular fiction, Parliament proceedings, and other sources. It is available online⁵⁹ free of charge along with a useful set of search commands. Since Polish is a highly inflected language, the corpus allows searching for individual inflected forms as well as for lemmas with all their inflected forms. All items are displayed along with their immediate context and a detailed grammatical coding. As much as the IPI PAN database is possibly the best Polish corpus available at the moment, it has some minor problems such as e.g. coding errors. Some common coding mistakes include the coding of proper names, e.g. “Macko” or “Zadro”, as vocative forms of nouns. Other mistakes include nouns coded as verbs, e.g. the word “para” (which means “a couple” or “two”) has been coded as the 3rd person singular of the imperfective verb “parać się”/ “occupy oneself with”: “Były jak **para** [parać:sg:3rd:imperf] wypalonych gwiazd” / “They resembled **two** burned-out stars”. Therefore, we found it necessary to check the coding of each occurrence of the relevant target words. In order to calculate the total frequency of occurrence of the target words, IPI PAN search results were pasted into Excel. Wrongly coded items were then deleted and total frequencies were calculated.

Polish verbs have two aspects (imperfective and perfective), four tenses (Present, Past, Compound Future and Simple Future) and two additional modes (conditional and imperative). Each verb is inflected for person (first person, second person, third person) in the singular and in the plural. Also a gender distinction is made in some tenses according to whether the verb forms are masculine, feminine or neuter. Gender occurs with the past, compound future and conditional tenses. Nonfinite verb forms are not inflected for person or number, e.g. the infinitive “zanurzyć” (“dip” e.g. hands in the water) or its adverbial participle “zanurzając”. The frequency of occurrence of the verbs included in the L1 Vocabulary Test ranges from 22 to 549 (see Tables A.10 and A.11 in Appendix H for detailed lists). Among the verbs from our

⁵⁸ The PWM Polish corpus can be accessed at <http://korpus.pwn.pl>.

⁵⁹ The IPI PAN corpus can be accessed at <http://korpus.pl/>.

pool of target verbs the word “zanurzyć” has the highest number of inflected forms in the IPI PAN corpus, namely 23 forms (see Table 6.9 below).

Table 6.9. Frequencies of occurrence of inflected forms of a selected Polish verb in the IPI PAN corpus

Base form	Inflected forms	Frequencies of inflected forms	Overall frequency	No on the test
zanurzyć / zanurzać (<i>dip</i>) [perf /imperf] ⁶⁰	zanurza [sg:3rd:imperf]	53	464	18
	zanurzać [inf:imperf]	26		
	zanurzacie [pl:2nd:imperf]	1		
	zanurzaj [impt:sg:2nd:imperf]	1		
	zanurzają [pl:3rd:imperf]	14		
	zanurzając [advpart:imperf]	23		
	zanurzał [sg:m:imperf]	20		
	zanurzała [sg:f:imperf]	11		
	zanurzali [pl:m:imperf]	9		
	zanurzało [sg:n:imperf]	3		
	zanurzały [pl:f:imperf]	6		
	zanurzam [sg:1st:imperf]	10		
	zanurzamy [pl:1st:imperf]	15		
	zanurzasz [sg:2nd:imperf]	2		
	zanurz [impt:sg:2nd:perf]	4		
	zanurzą [pl:3rd:perf]	8		
	zanurzcie [impt:pl:2nd:perf]	1		
	zanurzę [sg:1st:perf]	1		
	zanurzmy [impt:pl:1st:perf]	4		
	zanurzy [sg:3rd:perf]	12		
	zanurzyć [inf:perf]	121		
	zanurzycie [pl:2nd:perf]	2		
	zanurzył [sg:m:perf]	68		
	zanurzyła [sg:f:perf]	18		
	zanurzyli [pl:m:perf]	16		
	zanurzyło [sg:n:perf]	6		
	zanurzyły [pl:f:perf]	3		
	zanurzymy [pl:1st:perf]	6		

As mentioned above, the test also probes the knowledge of Polish nouns and adjectives. Polish adjectives agree with the nouns they modify in gender (masculine /feminine / neuter), number (singular / plural) and case. The following example – “piękna dziewczyna” / “beautiful girl” – illustrates seven cases that Polish nouns and adjectives take:

Nominative (nom)	piękna dziewczyna
Genitive (gen)	pięknej dziewczyny
Dative (dat)	pięknej dziewczynie
Accusative (acc)	piękną dziewczynę
Instrumental (inst)	piękną dziewczyną
Locative (loc)	pięknej dziewczynie
Vocative (voc)	piękna dziewczyno

⁶⁰ perf, imperf – perfective, imperfective (aspect)
1st, 2nd, 3rd – first, second, third (person)

sg, pl – singular, plural (number)
impt – imperative (mode)

The selected nouns and adjectives occur in the IPI PAN corpus less than 571 times and less than 316 times respectively (Appendix H). Within the pool of our target words the noun “parobek” / “farmhand” and the adjective “opieszwały” / “sluggish” have the highest number of inflections in the IPI PAN corpus (see Table 6.10 below).

Table 6.10. Frequencies of occurrence of inflected forms of a selected noun and adjective in the IPI PAN corpus

Base form	Inflected forms	Frequencies of inflected forms	Overall frequency	No on the test
parobek (noun) (farmhand)	parobcy [pl:nom:m] ⁶¹	41	387	47
	parobek [sg:nom:m]	89		
	parobka [sg:acc:m][sg:gen:m]	64		
	parobkach [pl:loc:m]	2		
	parobkami [pl:inst:m]	23		
	parobki [pl:nom:m]	23		
	parobkiem [sg:inst:m]	23		
	parobkom [pl:dat:m]	9		
	parobków [pl:acc:m][pl:gen:m]	92		
	parobkowi [sg:dat:m]	8		
	parobku [sg:loc:m][sg:voc:m]	13		
opieszwały (adj) (sluggish)	opieszwał [sg:acc:f] [sg:inst:f]	7	146	1
	opieszwała [sg:nom:f]	13		
	opieszwał [pl:acc:f][pl:acc:n]			
	[pl:nom:f][pl:nom:m]	33		
	opieszwałego [sg:acc:m][sg:acc:n]	21		
	opieszwałej [sg:gen:f]	2		
	opieszwałemu [sg:dat:m]	1		
	opieszwali [pl:nom:m]	6		
	opieszwały [sg:acc:m][sg:nom:m]	28		
	opieszwałych [pl:acc:m]			
	[pl:gen:m][pl:loc:m]	22		
	opieszwałym [pl:dat:m][pl:inst:m]			
	[pl:inst:n]	10		
	opieszwałymi [pl:inst:m][pl:inst:n]	3		

The number of letters provided to prompt each target word ranges from 2 to 4 letters and it never exceeds half of the target word stem, as illustrated in the following example:

Osoby, mieszkające w dużych aglomeracjach, są czasami pogardliwie nazywane **mie**_____. (*mieszczuchami*)

People living in large urban agglomerations are sometimes called disapprovingly *city-slickers*.

At the piloting stage as well as at the main experiment the test items were presented on a computer screen in randomized order. The subjects were given up to 40 minutes to complete the test. The measure of performance on the L1 Vocabulary Test

⁶¹ sg, pl – singular, plural (number)
m, f, n – masculine, feminine and neuter (gender)
nom, gen, dat, acc, inst, loc, voc – nominative, genitive, dative, accusative, instrumental, locative and vocative (cases)

was the overall score obtained by each student. The test was scored on a dichotomous scale (0,1) and minor spelling mistakes were disregarded. The maximum possible score was 54.

6.3.8. OUP Quick Placement Test

The OUP Quick Placement Test is a test of English language proficiency available in two parallel versions (Appendix I). Its manual (OUP, 2001) includes guidelines on how to interpret the obtained results in relation to the CEF levels (presented in length in Chapter 4). All the questions in the test are in a multiple-choice format. The paper and pen version consists of two parts assessing reading, vocabulary and grammar. Part 1 is meant to be taken by all students irrespective of their level of proficiency. Part 2 was devised for higher ability students only (upper-intermediate and higher). The test takes up to 30 minutes to administer, and is relatively quick to check.

The test designers strongly recommend that the test be used alongside another form of assessment, as its standard error of measurement is + or – 4 points on the 60-point scale (OUP, 2001). Therefore, an additional writing task was devised and added as the last section of the test (see Appendix I). The results obtained in the Quick Placement Test were cross-referenced with the level and complexity of student writing.

Chapter 7: Results and Discussion

7.1. Introduction

The present chapter comprises seven sections apart from this introductory section, four of which address the research questions outlined in Chapter 1 (Section 7.5-7.8). The remaining three sections contain relevant background information, which is critical to the interpretation of the findings and to the confidence we can place in our analyses and conclusions – for example, the analysis of the assumptions of parametric data (Sections 7.2-7.4).

The chapter begins with an overview of the measures and scoring methods employed in the study (a detailed presentation of the research instruments can be found in Chapters 5 and 6). Section 7.2, apart from listing the research instruments, also contains guidelines as to where in the present chapter each of the research questions is addressed. Section 7.3 analyses whether the data satisfy the assumptions of normality and homogeneity for the use of parametric tests and presents initial results. In the interests of clarity and simplicity, the distribution of each set of scores is presented graphically via boxplots (means and standard deviations included), and relevant numerical data are placed in Tables A.16 – A.27 in Appendix K for further inspection. Section 7.4 presents half-split reliability coefficients for our language and memory measures, and it addresses the issue of the construct validity of our instruments.

Section 7.5 focuses on the correlational links among our variables. It looks at the strength of correlations among the variables classified in the study as independent variables in order to test the assumption of no multicollinearity. It discusses the strength of the correlations between the independent and dependent variables in the light of our research questions, and relates our findings to the existing literature. Sections 7.6 and 7.7 continue the discussion of these research questions with the use of multiple linear regression models. Sections 7.5 – 7.7 make no direct inferences about the causal direction between our variables; the issue of causality is explicitly addressed in Section 7.8 via the application of a cross-lagged panel correlation paradigm.

7.2. Brief overview of measures and research questions

The study involves a number of independent variables, namely Serial Nonword Recall (tapping the phonological store and the articulatory component of PSTM), Serial Nonword Recognition (tapping the phonological store), exposure to L2 outside the classroom, and L1 (low-frequency) vocabulary knowledge (see Table 7.1 below). The dependent variables include the following: L2 vocabulary learning/knowledge outside the classroom, L2 vocabulary learning in class, L2 collocation learning/knowledge outside the classroom, and L2 collocation learning in class (see Chapter 1, Section 1.2. for a description of each of the constructs). Please note that in Section 7.8 some of the dependent variables become independent variables and vice versa. Scoring methods for each variable are summarised in Table 7.1 below.

Table 7.1. Overview of variables, measures and scoring methods

Type of variable	Variable name	Corresponding test scores	Scoring method
Independent variables	PSTM capacity (the phonological store and the articulatory component)	Serial Nonword Recall scores	1 point for each correctly repeated list of nonwords (correct-in-position criterion)
	PSTM capacity (the phonological store)	Serial Nonword Recognition scores	1 point for each correctly recognised set of lists of nonwords (as the same or different)
	Exposure to L2 outside the classroom	number of hours per day recorded in students' journals ^A	averaged number of hours per day
	L1 (low-frequency) vocabulary knowledge (productive)	L1 Vocabulary Test scores	1 point for each correctly produced word
Dependent variables	L2 vocabulary learning/knowledge outside the classroom (receptive and productive)	change in A2 and B1 Vocabulary Levels Test scores A2 and B1 Vocabulary Levels Test scores (Time 1 and 2)	Scores at Time 1 subtracted from scores at Time 2 1 point for each correctly produced or correctly recognised L2 word
	L2 vocabulary learning in class (receptive and productive)	A2 and B1 Progress Vocabulary Test scores	1 point for each correctly produced or correctly recognised L2 word
	L2 collocation learning/knowledge outside the classroom (productive)	change in A2 and B1 Gitsaki Collocation Test scores A2 and B1 Gitsaki Collocation Test scores (Time 1 and 2)	Scores at Time 1 subtracted from scores at Time 2 1 point for each correctly produced or correctly recognised L2 collocation
	L2 collocation learning in class (receptive and productive)	A2 and B1 Progress Collocation Test scores	1 point for each correctly produced or correctly recognised L2 collocation

Note: Minor spelling mistakes were disregarded.

^A Information on exposure to L2 outside the classroom was collected via students' journals between Time 1 (January 2008) and Time 2 (July 2008) in selected six weeks (one week from each month).

The present Ph.D. thesis investigates the relationship between L1-based (Polish) Phonological Short-term Memory measures (the Serial Nonword Recall task and the Serial Nonword Recognition task⁶²) and L2 (English) vocabulary and collocation learning in a longitudinal classroom-based study involving young adult native speakers of Polish resident in an English-speaking country (Ireland). Three of our research questions (Questions 1, 2 and 3 in Table 7.2 below) examine the link between PSTM and the development of L2 vocabulary knowledge in adult learners, while the remaining three (Questions 4, 5 and 6) look at the link between PSTM and the development of L2 collocational knowledge in adult learners at the A2 and B1 levels of proficiency. The methods used to explore the link between the above mentioned constructs are simple correlations (Section 7.5), simple linear regressions and hierarchical multiple regressions (Sections 7.6), and cross-lagged correlations (Section 7.2). The sections in which each of our research questions is addressed are specified in Table 7.2 below.

Table 7.2. Overview of research questions and sections where the questions are addressed

Research questions		Sections
Question 1	Does the relationship between PSTM and L2 vocabulary learning (in class and outside the classroom) decrease with an increased level of L2 proficiency in adults, as has been shown to happen in the case of child L1 learning?	Sections 7.5 and 7.6
Question 2	Which component of PSTM is more closely linked to L2 vocabulary learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?	Sections 7.5 and 7.6
Question 3	In the case of adults, what is the causal direction of the relationship between PSTM and the learning of high-frequency L2 vocabulary (outside the classroom) at the A2 and B1 levels of L2 proficiency? Is it the L1-based nonword scores at Time 1 that contribute to L2 Vocabulary Levels Test scores at Time 2, or is it L2 Vocabulary Levels Test scores at Time 1 that contribute to the L1-based nonword scores at Time 2?	Section 7.8
Question 4	In the case of adults, is there a relationship between PSTM and the learning of L2 collocations (in class and outside the classroom) at the A2 and B1 levels of L2 proficiency? If so, does the relationship between PSTM and L2 collocation learning decrease with an increased level of L2 proficiency in adults?	Sections 7.5
Question 5	Which component of PSTM is more closely linked to L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?	Sections 7.5 and 7.7
Question 6	What is the causal direction of the relationship between PSTM and L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency? Is it PSTM scores at Time 1 that contribute to Gitsaki Collocation Test scores at Time 2, or is it Gitsaki Collocation Test scores at Time 1 that contribute to the PSTM scores at Time 2?	Section 7.8

Note: Time 1 – January 2008; Time 2 – July 2008.

⁶² Please note that for ease of reading the names of PSTM tasks are capitalised throughout.

7.3. Distribution of scores and assumptions of data

A failure to properly check data assumptions in quantitatively-based research projects can lead to the use of inappropriate statistical procedures and, as a result, flaw the logic behind hypothesis testing (Field, 2005). There are four basic assumptions of parametric data, namely the assumptions of normality, homogeneity of variance, interval data and data independence. The first two assumptions can be tested graphically and numerically, and will be dealt with in great depth in this chapter. The remaining two assumptions can be tested only by common sense (Field, 2005), and hence no numerical data will be presented to support them.

The assumption of normality is met when data are from one or more normally distributed populations. The data can be assessed by eye using histograms to assess whether it looks roughly normal, but it should also be subjected to other, more stringent analyses. A good starting point is to generate boxplots to spot outliers and some obvious mistakes made during data entering. The next step is to look at numerical data that describe the shape of the distribution, such as skewness and kurtosis. Skewness characterises the degree of asymmetry of a distribution around the mean, while kurtosis characterises the relative degree of flatness or peakedness of the distribution as compared to the normal distribution. Normal distributions produce skewness and kurtosis statistics of around zero. Many researchers use the criterion that kurtosis and skewness values should be within the +2 to -2 range, when data are normally distributed. Some authors use +1 to -1 as a more stringent criterion. In smaller samples like ours, the value of skewness and kurtosis should be transformed to a z-score to calculate the significance of skewness and kurtosis, as the values on their own are not very informative. When the skewness and kurtosis statistics are non-significant, it means that they are not statistically different from zero. The final step is to run normality tests (e.g. the Kolmogorov-Smirnov or Shapiro-Wilk tests⁶³) and generate normality plots (e.g. P-P or Q-Q plots).

⁶³ The Kolmogorov-Smirnov and Shapiro-Wilk tests are formal tests of normality offered in statistical software such as SPSS. Owing to the fact that the two tests have different limitations (the discussion of which is beyond the scope of this thesis) we decided to use both to double-check that scores under scrutiny are not significantly different from a normally distributed set of scores with the same mean and standard deviation.

The assumption of homogeneity of variance is met when the variance within each sample of data is equal. Variance is a measure of the average distance between each of a set of data points and their mean value. In designs in which several groups of participants are tested this assumption implies that each sample of data comes from populations with the same variance. In correlational studies, the assumption of homogeneity of variance means that the variance of one variable is assumed to be the same at all levels of the other variable. The test that checks the assumption of homogeneity of variance in the statistical software used in the current study, SPSS 16.0, is called Levene's test. When the Levene statistic is non-significant, the variances are roughly equal and the assumption of homogeneity of variance is tenable.

Our analysis of assumptions of parametric data involves both graphic and numerical methods. Each distribution was first assessed by eye using histograms and boxplots. Next, skewness and kurtosis statistics were standardized in order to establish whether they were statistically different from zero. Finally, the Kolmogorov-Smirnov or Shapiro-Wilk tests were carried out, and normality plots were generated. Boxplots are included in this section to complement the presentation of descriptive statistics of each variable. Detailed tables with descriptive statistics, standard deviation values, skewness and kurtosis values, normality tests' statistics and the Levene statistics are presented in Appendix K. Selected normality plots are also presented in Appendix K.

Serial Nonword Recall and Recognition scores (Time 1, Time 2)

The analysis of A2 Serial Nonword Recall, A2 Serial Nonword Recognition, B1 Serial Nonword Recall, and B1 Serial Nonword Recognition scores ($n=30$ in each case) reveals that the four sets of scores are normally distributed at Time 1. The Kolmogorov-Smirnov and the Shapiro-Wilk tests indicate no significant deviations from normality (all $ps > .05$; see Appendix K, Table A.16). When transformed to a z -score, the values of skewness and kurtosis in the four sets of data are non-significant (all $ps > .05$), which confirms the assumption that the sets of scores are normally distributed. The generated normality plots show that the distribution of the variables is normal, as points cluster around the straight line in each of the four plots (see Appendix K, Figures A.9 to A.12). Boxplots of Serial Nonword Recall and Recognition scores in the A2 and B1 groups do not reveal any outliers in the data (Figure 7.1 below). On the basis of the above we conclude that the assumption of normality of PSTM scores in all

four groups is tenable.

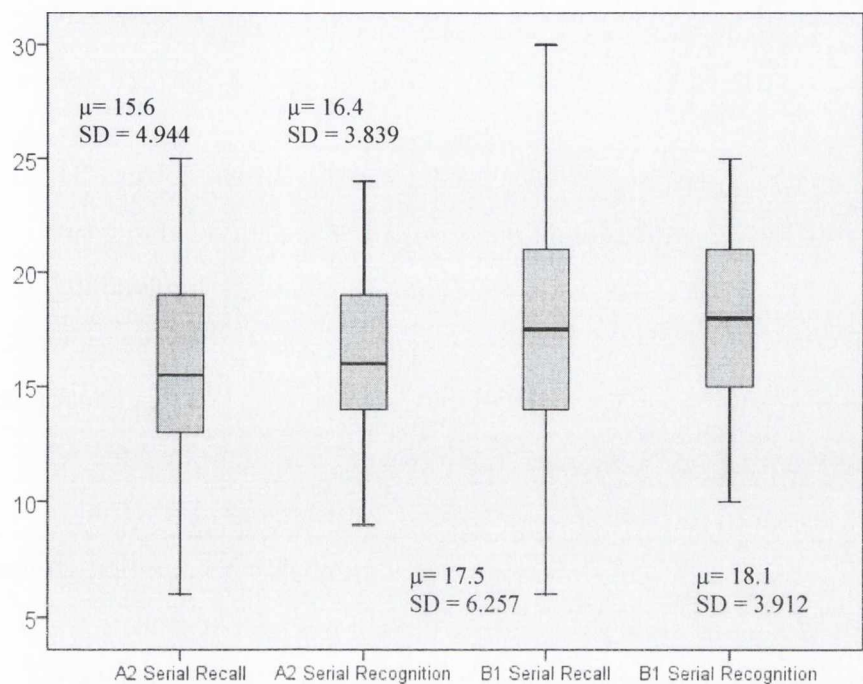


Figure 7.1. Boxplots of Serial Nonword Recall and Recognition scores in A2 and B1 groups (Time 1)

The lowest score obtained in the A2 Serial Nonword Recall group and that obtained in the B1 Serial Nonword Recall group were the same at Time 1, namely 6 correctly repeated lists. The highest scores in the A2 Serial Nonword Recall group and the B1 Serial Nonword Recall group were 25 and 30 lists respectively. The scores in the A2 Serial Nonword Recognition group and the B1 Serial Nonword Recognition group ranged from 9 to 24 lists and from 10 to 25 lists respectively. Boxplots in Figure 7.1 above illustrate visually that the range of A2 Serial Nonword Recall scores is narrower than it is in the B1 Serial Nonword Recall group. The range of A2 Serial Nonword Recognition scores and that of A2 Serial Nonword Recognition scores are the same.

The means of scores in the four groups appear to indicate that the B1 students perform better than the A2 students on both PSTM tasks (see Appendix K, Table A.16). The means of A2 and B1 Serial Nonword Recall scores, however, are not statistically different ($t(58) = -1.30, p > .05$), and neither are the means of A2 and B1 Serial Nonword Recognition scores ($t(58) = -1.63, p > .05$). A2 Serial Nonword Recall and Recognition scores correlate at the .001 level ($r = .683$), and so do B1 Serial Nonword Recall and Recognition scores ($r = .645$). The strength of these correlations resembles

that obtained during the pilot study (Chapter 5, Section 5.3.1). When corrected for attenuation (Chapter 5, Equation 5.1), the correlation values are .782 and .812 respectively.

In order to compare scores obtained by students with higher and lower PSTM scores, A2 and B1 Serial Recall and Recognition scores were each divided into two halves according to the level of PSTM functioning. Each High PSTM group contains students with the top 15 scores on Serial Nonword Recall and Recognition tasks, while each Low PSTM group contains students with the lowest 15 scores. The resultant eight PSTM groups include the A2 High and Low Serial Nonword Recall groups, the A2 High and Low Serial Nonword Recognition groups, the B1 High and Low Serial Nonword Recall groups and the B1 High and Low Serial Nonword Recognition groups. The distributions of scores in the eight PSTM groups conform to the normal distribution (see normality tests' statistics, Levene's test statistics, and skewness and kurtosis values in Tables A.17 and A.18). Independent samples t-test procedures show that the means of High and Low PSTM groups are significantly different (for A2 High and Low Serial Nonword Recall scores $t(28) = -7.864$, $p < .001$; for A2 High and Low Serial Nonword Recognition scores $t(28) = -8.159$, $p < .001$; for B1 High and Low Serial Nonword Recall scores $t(28) = -8.011$, $p < .001$; for B1 High and Low Serial Nonword Recognition scores $t(28) = -7.933$, $p < .001$). This indicates that scores in High and Low PSTM groups are well enough spread to distinguish between high and low PSTM capacity students.

Subjects' performance on the Serial Nonword Recall and the Serial Nonword Recognition tasks was measured again at Time 2 (i.e. the end of the TCD language course), resulting in four additional variables – Serial Nonword Recall (A2), Serial Nonword Recall (B1), Serial Nonword Recognition (A2) and Serial Nonword Recognition (B1). The range of the scores of the four variables obtained at Time 2 resembles that obtained at Time 1. The range of scores is presented graphically in Figure 7.2 below and numerically in Table A.19 (Appendix K). The four sets of scores are normally distributed (see Table A.19, Appendix K for relevant statistics). The boxplot of the Serial Recognition scores in the B1 group (Figure 7.2 below) indicates the presence of a possible outlier (case 24). The score of student 24 was not a scoring error (the data was collected via a dedicated website and scored electronically in

Excel), and therefore we decided against excluding case 24 from further analyses. The difference between group means at Time 1 and Time 2 were non-significant for each pair of the measures, which implies that group means did not change significantly between Time 1 and Time 2 ($t(29) = 0.290$, $p > .05$ for the A2 Serial Nonword Recall at Time 1 and Time 2; $t(29) = 0.976$ for the A2 Serial Nonword Recognition at Time 1 and Time 2; $t(29) = 0.094$ for the B1 Serial Nonword Recall at Time 1 and Time 2; $t(29) = -0.088$ for the B1 Serial Nonword Recognition at Time 1 and Time 2).

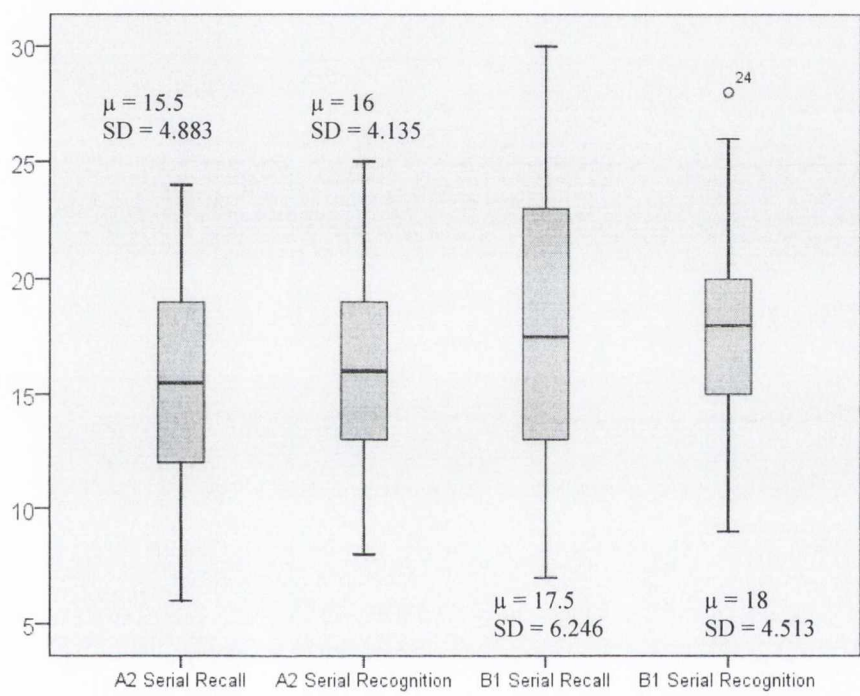


Figure 7.2. Boxplots of Serial Nonword Recall and Recognition scores in A2 and B1 groups (Time 2)

L1 Vocabulary Test scores (Time 1)

The distributions of A2 and B1 L1 Vocabulary Test scores do not differ significantly from the normal distribution. Normality plots (Appendix K, Figures A.16 and A.17) and boxplots (Figure 7.3 below) indicate that the data are normally distributed and that there are no outliers. The Kolmogorov-Smirnov test statistics and the Shapiro-Wilk test statistics are non-significant in both groups ($ps > .05$). The values of skewness and kurtosis in the A2 and B1 distributions are not significantly different from zero (see Appendix K, Table A.20). As can be seen in Figure 7.3 below, the L1 Vocabulary Test scores are slightly more spread out in the A2 group than they are in the B1 group. The highest scores obtained in the A2 and B1 groups were 98.1 and 96.3

per cent respectively. The lowest A2 and B1 scores were 61.1 and 64.8 per cent respectively. The mean of the A2 students' scores was 78.1 per cent, and it was 3.7 per cent lower than the mean of the B1 students' scores (Appendix K, Table A.20). A t-test for equality of means indicates that the difference between the two means is non-significant ($t(58) = -1.461, p > .05$).

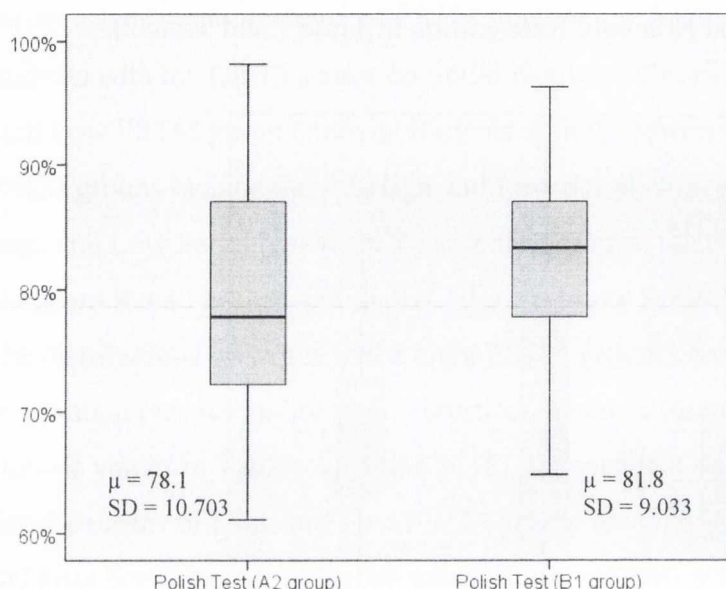


Figure 7.3. Boxplots of L1 Vocabulary Test scores in A2 and B1 groups (Time 1)

Exposure to L2 outside the classroom (between Time 1 and Time 2)

Our analysis shows that Exposure scores in the A2 and B1 groups are normally distributed. The Kolmogorov-Smirnov test statistics and the Shapiro-Wilk test statistics indicate that the two distributions are not significantly different from the normal distribution (all p s $> .05$). The values of skewness and kurtosis in the A2 and B1 distributions are not significantly different from zero, as expected in normally distributed data (see Appendix K, Table A.21). Boxplots in Figure 7.4 below do not indicate any outliers. Points cluster around a straight line in normality plots. Levene's test confirms the assumption of homogeneity of variance ($F(1,28) = 0.018, p > .05$ in the A2 group, $F(1,28) = 0.108, p > .05$ in the B1 group). Exposure scores in the A2 and B1 groups ranged from one to eight and two to nine hours per day respectively. The mean Exposure score in the A2 group was 4.4 hours per day, while the mean Exposure score in the B1 group was 5.4 hours per day. The B1 group appears to have had one

hour more of Exposure per day than the A2 group did, but the difference between the group means is not significantly different from zero ($t(58) = -1.74, p > .05$).

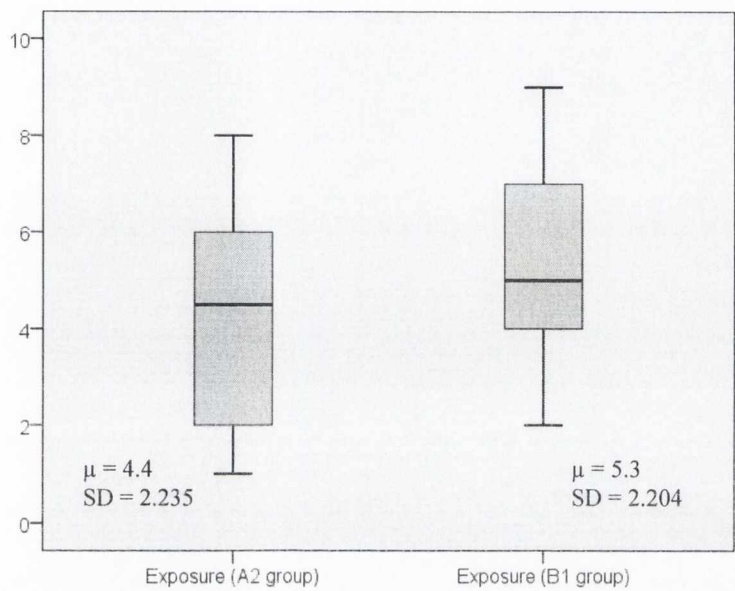


Figure 7.4. Boxplots of Exposure scores in A2 and B1 groups

A2 and B1 Vocabulary Levels Test scores (Time 1, Time 2)

The distributions of the A2 Vocabulary Levels Test scores at Time 1 and Time 2 are not significantly different from the normal distribution. The Kolmogorov-Smirnov test and the Shapiro-Wilk test indicate no significant differences from normality ($ps > .05$). Values of skewness and kurtosis are within the +2 to -2 range, and are not significantly different from zero (all $ps > .05$). According to Levene's test the assumption of homogeneity of variance is tenable both at Time 1 and Time 2 (see Appendix K, Table A.22). The A2 Vocabulary Levels Test scores at Time 1 ranged from 30.4 per cent to 70.3 per cent (see Figure 7.5 below). The minimum score obtained at Time 2 was over one per cent higher than that obtained at Time 1, namely 31.7 per cent, while the maximum score was over 10 per cent higher than that at Time 1 reaching 80.9 per cent. The difference between Time 1 mean (49.2 per cent) and Time 2 mean (56.1 per cent) is highly significant ($t(29) = 9.243, p < .001$).

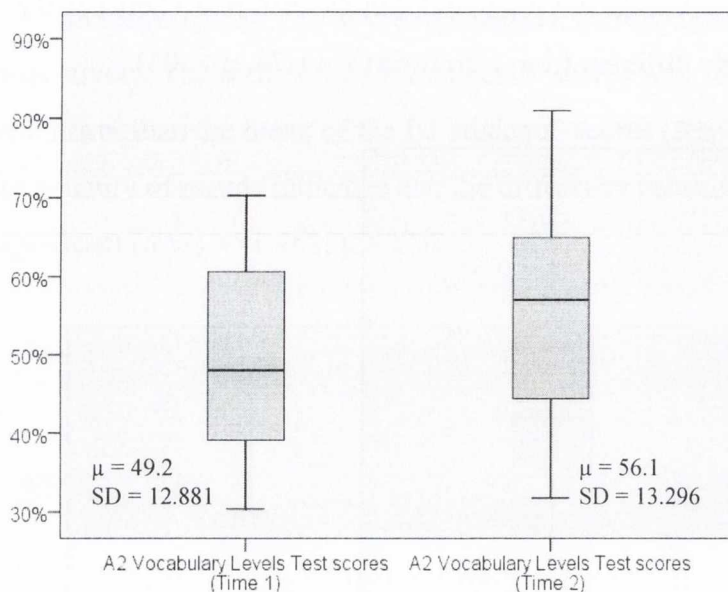


Figure 7.5. Boxplots of A2 Vocabulary Levels Test scores (Time 1 and Time 2)

A2 Vocabulary Levels Test gain scores between Time 1 and Time 2 also normally distributed. As can be seen in Figure 7.6 below, the distribution is positively skewed (0.623), but the value of skewness is non-significant ($p > .05$). The value of kurtosis is very close to zero (0.030) and is also non-significant ($p > .05$). The statistics of the Kolmogorov-Smirnov test, the Shapiro-Wilk test and Levene's test are all non-significant ($ps > .05$). The gains in A2 Vocabulary Levels Tests our A2 subjects obtained over the period of 6 months ranged between 0.7 and 16.2 per cent with the mean of 6.9 per cent.

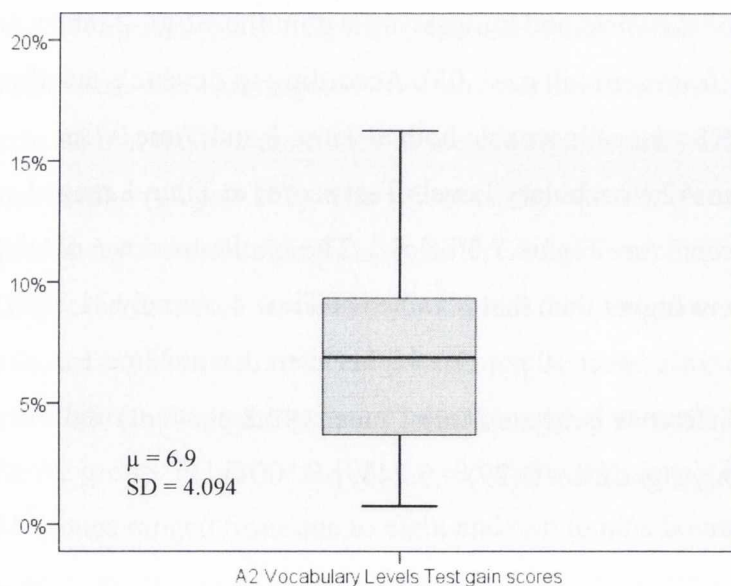


Figure 7.6. Boxplot of gains in A2 Vocabulary Levels Test scores

The scores on B1 Vocabulary Levels Test are normally distributed at Time 1 and Time 2. The Kolmogorov-Smirnov test and the Shapiro-Wilk test indicate no significant deviations from the normal distribution (all p s > .05). The Levene statistics are in both cases non-significant ($F(1,28) = 3.645$ for Time 1 scores and $F(1,28) = 1.956$ for Time 2 scores). Time 1 and Time 2 distributions are negatively skewed (skewness of -0.660 for Time 1 scores and -0.484 for Time 2 scores) and they are less peaked than the normal distribution (kurtosis of -0.111 and -0.344 for Time 1 scores and Time 2 scores respectively). The values of skewness and kurtosis are not significantly different from zero (see Appendix K, Table A.23).

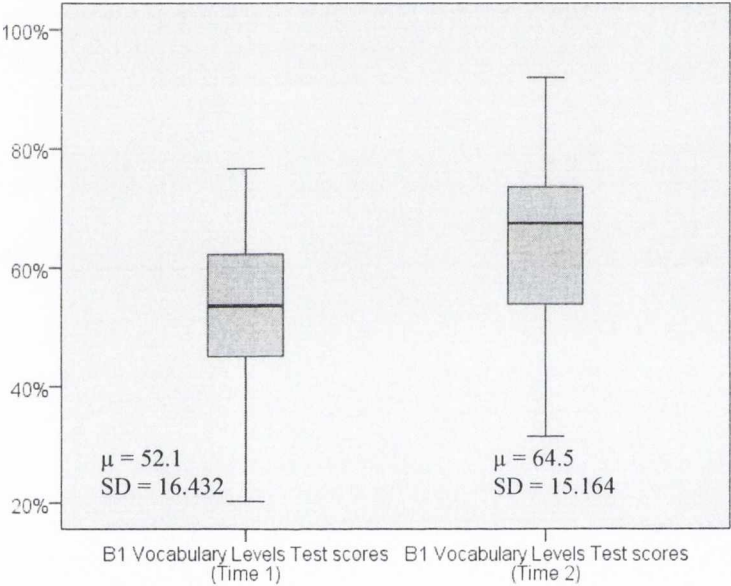


Figure 7.7. Boxplots of B1 Vocabulary Levels Test scores (Time 1 and Time 2)

The scores on the B1 Vocabulary Levels Test ranged from 20.3 per cent and 76.7 per cent at Time 1 and from 31.3 to 91.9 per cent at Time 2 (see Figure 7.7 above). The B1 students gained from 1.1 to 26.3 per cent on the B1 Vocabulary Levels Test over the period of six months. The B1 Vocabulary Levels Test gain scores, represented graphically in Figure 7.8 below, conform to the normal distribution (see Appendix K, Table A.24). The difference between the group mean at Time 1 and that at Time 2 is highly significant ($t(29) = 9.858$, $p < .001$). Interestingly, the difference between the minimum score at Time 1 and that at Time 2 turned out to be rather high as compared to that on the A2 Vocabulary Levels Test (11 and 1.3 per cent respectively). In order to ensure that the minimum score at Time 1 was not a result of a coding error, we double-checked the Excel conditional formulas used to calculate B1 Vocabulary Levels Test

scores. No formula errors were identified. We also reviewed the subject's OUP proficiency test scores and the writing task that was administered along with the OUP test. Since the minimum score at Time 1 was not a result of a coding error and the subject was assigned to the right proficiency group, we did not exclude that case from our further analyses.

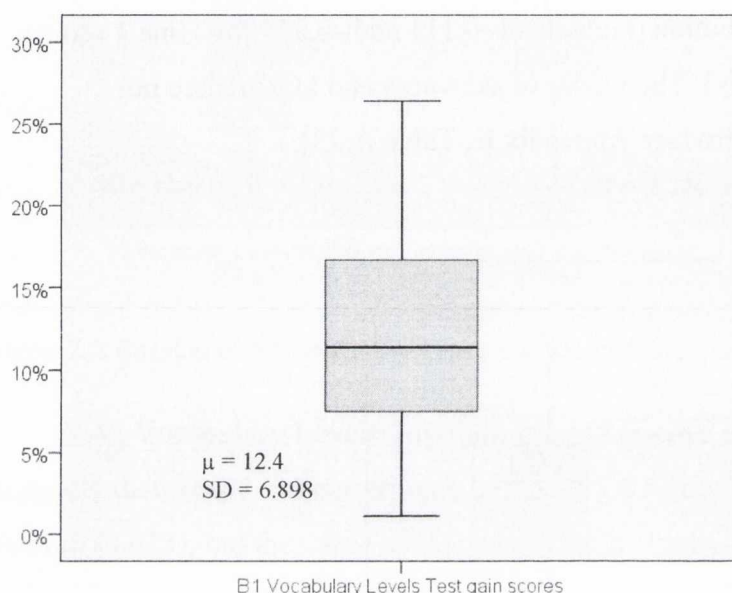


Figure 7.8. Boxplot of gains in B1 Vocabulary Levels Test scores

A2 and B1 Gitsaki Collocation Test scores (Time 1, Time 2)

The A2 Gitsaki Collocation Test scores at Time 1 and Time 2 are normally distributed (see Figure 7.9 below). Normality tests are non-significant for both distributions of scores ($p > .05$). The values of skewness and kurtosis are within the +1 to -1 range and are also non-significant (see Appendix K, Table A.25). The assumption of homogeneity of variance is supported by the obtained Levene statistic ($F(1,28) = 0.893, p > .05$). The A2 Gitsaki Collocation Test scores at Time 1 and Time 2 ranged between 4 and 54 per cent and between 6 and 62 percent respectively. The group mean at Time 1 was 27.03 per cent, while the group mean at Time 2 was 33.9 per cent. The difference between the two means is statistically significant ($t(29) = 7.210, p < .001$).

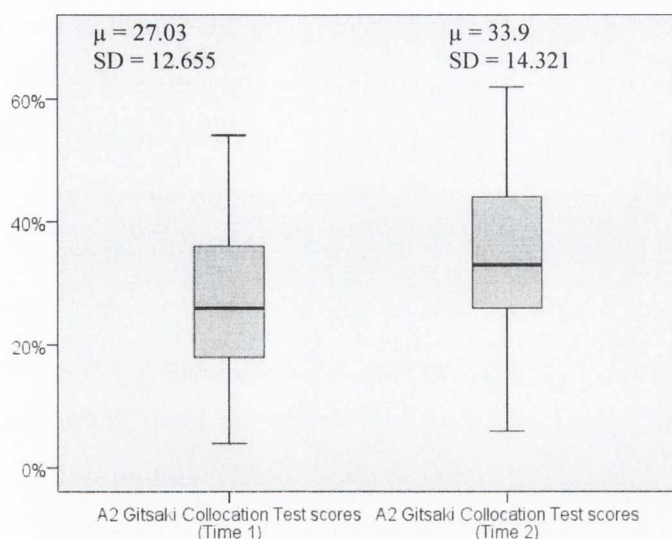


Figure 7.9. Boxplots of A2 Gitsaki Collocation Test scores (Time 1 and Time 2)

The A2 Gitsaki Collocation Test gain scores are normally distributed (see Figure 7.10 below). The Kolmogorov-Smirnov test and the Shapiro-Wilk test statistics are non-significant ($p > .05$, see Appendix K, Table A.X). The values of skewness and kurtosis (0.306 and -1.196 respectively) are not significantly different from zero. The probability plot also suggests normally distributed data. The assumption of homogeneity of variance is tenable ($F(1,28) = 0.757$, $p > .05$). Over the period of six months between Time 1 and Time 2 the group of A2 students showed gains of 6.9 per cent on average. The gains on the A2 Gitsaki Collocation Test ranged between zero and 16 per cent.

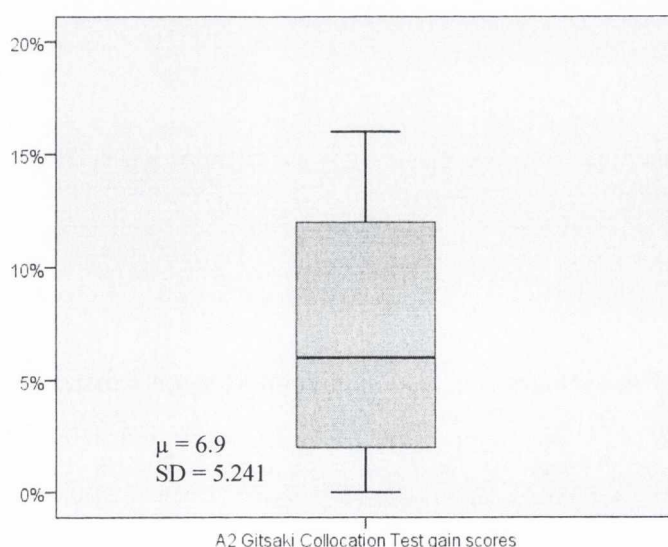


Figure 7.10. Boxplot of gains in A2 Gitsaki Collocation Test scores

The B1 Gitsaki Collocation Test scores (Time 1 and 2) and the B1 Gitsaki Collocation Test gain scores conform to the normal distribution. The values of skewness and kurtosis in the three sets of scores are non-significant when transformed to a z-score (all $ps > .05$; see Table A.26, Appendix K). The Kolmogorov-Smirnov statistic, the Shapiro-Wilk statistic and the Levene statistic are non-significant for all three distributions (all $ps > .05$). Scores at Time 1 range from 32 to 80 per cent. Scores at Time 2 range from 38 to 88 per cent. The mean score obtained at Time 1 was 55.03 per cent, while the mean score obtained at Time 2 was 63.1 per cent (the difference is significant at the .001 level, $t(29) = -15.625$). Between Time 1 and Time 2 our B1 subjects improved their scores on the B1 Gitsaki Collocation Test scores by 8.1 per cent on average.

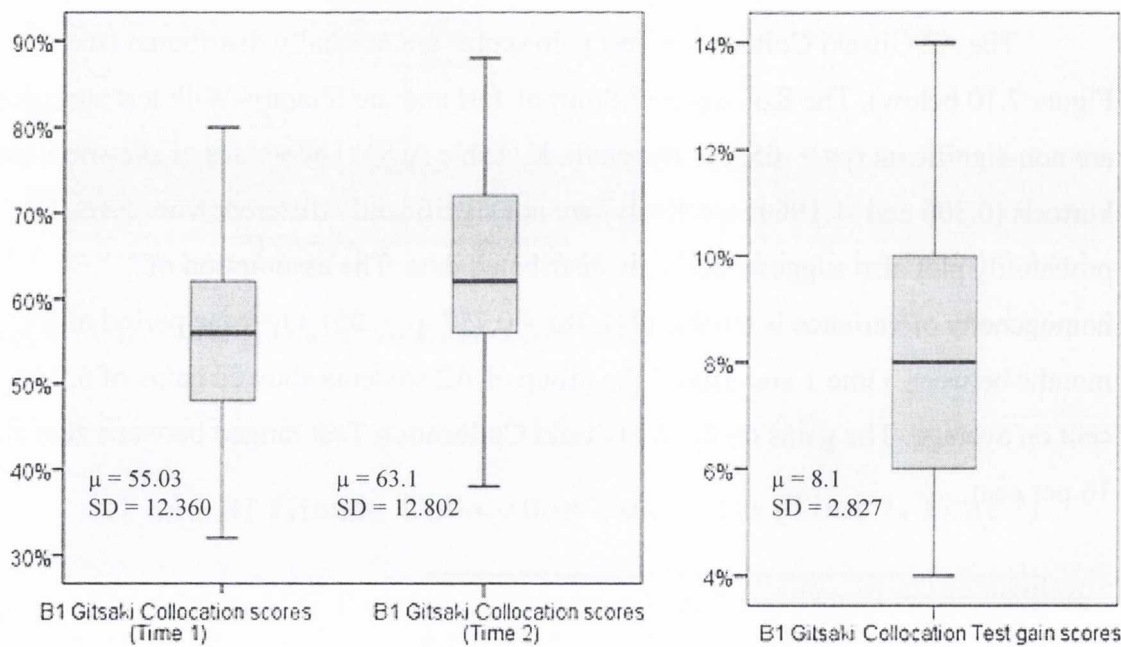


Figure 7.11. Boxplots of B1 Gitsaki Collocation Test scores (Time 1 and Time 2) and gains in B1 Gitsaki Collocation Test scores

A2 and B1 Progress Vocabulary Test scores (Time 2)

The distributions of the A2 and B1 Progress Vocabulary Test scores differ significantly from normal distribution. Boxplots in Figure 7.12 below indicate that both distributions have a noticeable negative skew. When the values of skewness and kurtosis are transformed to a z-score, the value of skewness in the A2 group is significantly different from zero ($p > .05$), while the remaining values are non-

significant ($ps < .05$). The Kolmogorov-Smirnov statistic, the Shapiro-Wilk statistic, and the Levene statistic are significant (all $ps < .05$, see Table A.28, Appendix K) for the A2 and B1 Progress Vocabulary Test scores, which indicates deviations from normality in both sets of scores. The A2 Progress Vocabulary Test scores range from 46.6 to 96.6 per cent, while the B1 Progress Vocabulary Test scores range from 53.3 to 96.7 per cent.

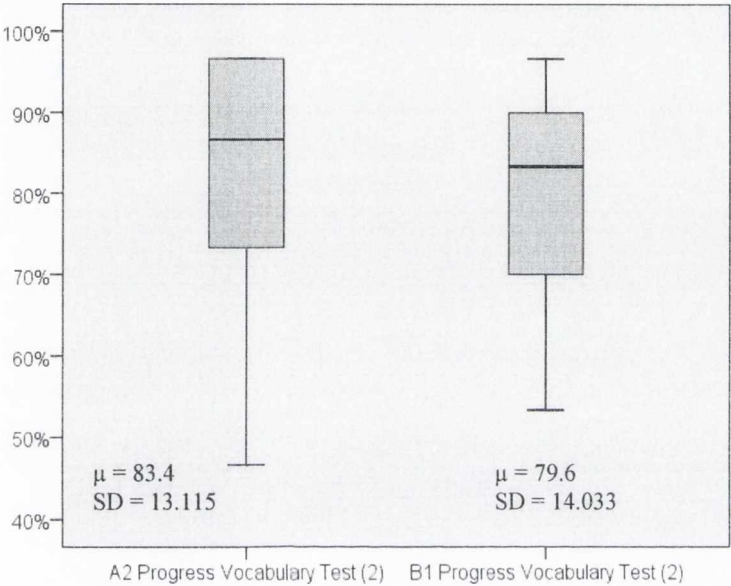


Figure 7.12. Boxplots of A2 Progress Vocabulary Test scores (Time 2) and B1 Progress Vocabulary Test scores (Time 2)

A2 and B1 Progress Collocation Test scores (Time 2)

The distributions of the A2 and B1 Progress Collocation Test scores also differ significantly from normal distribution. In both groups the values of skewness and kurtosis are not significantly different from zero ($p > .05$). However, the Shapiro-Wilk test indicates that both the A2 and B1 Progress Collocation Test scores are not normally distributed ($p < .05$ and $p < .01$ respectively)⁶⁴. Additionally, the Levene statistic is significant for the A2 Progress Collocation Test scores (see Table A.29, Appendix K). The A2 and B1 group means are 58.5 and 68.1 per cent respectively. The A2 Progress Collocation Test ranges from 46.6 to 78.8 per cent, while the B1 Progress

⁶⁴ The Shapiro-Wilk statistic for the A2 Progress Collocation Test scores is highly significant ($p < .01$), while the Kolmogorov-Smirnov statistic is non-significant ($p > .05$). Since the Shapiro-Wilk statistic is recommended for smaller samples like ours (while the Kolmogorov-Smirnov statistic is normally recommended for larger samples), we decided to treat the A2 Progress Collocation Test scores as nonparametric data.

Collocation Test ranges from 47.5 to 90.4 per cent (see Figure 7.13 below for a visual representation of the data).

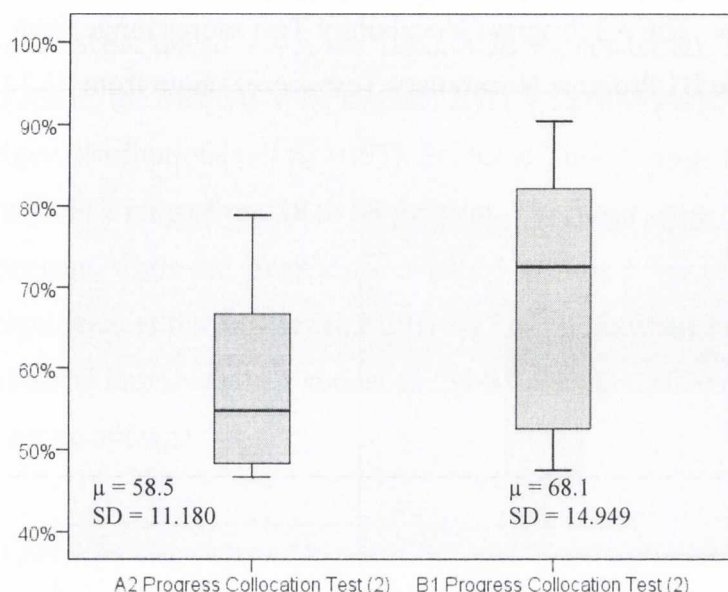


Figure 7.13. Boxplots of A2 Progress Collocation Test scores (Time 2) and B1 Progress Collocation Test scores (Time 2)

7.4. Reliability and validity of the research instruments

The first part of this section discusses the notion reliability of our research instruments, while the second part addresses the question of the instruments' validity. Reliability coefficients higher than .70, .80 and .90 are accepted here as representing adequate, good and very good internal consistency reliability respectively. With regard to validity, two types of construct validity are considered, namely discriminant validity and convergent validity.

The split-half reliability coefficients for Serial Nonword Recall at Time 1 (Spearman-Brown for unequal length) are very high in the A2 and B1 groups, .918 and .897 respectively (see Table 7.3 below for a list of Time 1 and Time 2 coefficients). The split-half reliability coefficients for Serial Nonword Recognition at Time 1 (Spearman-Brown for equal length) are satisfactory, but not nearly as good as the Serial Nonword Recall reliability coefficients. The split-half reliability coefficient calculated on the basis of the A2 Serial Nonword Recognition scores is comparable to that obtained during the pilot study (Chapter 5, Section 5.3.1), namely .831. The B1 split-half reliability coefficient is considerably lower and equals .704. The reliability

coefficients of the Serial Nonword Recognition measure are satisfactory. They are lower than we would ideally wish for, despite having adhered to guidelines that have been followed by other PSTM researchers using nonword tasks.

The L1 and L2 tests included in our test battery have satisfactory Spearman-Brown split-half reliability coefficients. The split-half reliability coefficients of the L1 Vocabulary Test at Time 1 are .876 and .864 for the A2 and B1 groups respectively and are comparable to those obtained during the pilot study. The test items were split into two halves according to part of speech and frequency of occurrence in the in the IPI PAN corpus (see Appendix H, Table A.11 for the IPI PAN frequencies of the Polish target words included in the L1 Vocabulary Test). The split-half reliability coefficients of the A2 and B1 Vocabulary Levels Tests at Time 1 are .876 and .951 respectively. The test items were split according to part of speech and frequency of occurrence in the BNC. The split-half reliability of the A2 and B1 Progress Vocabulary Test at Time 2 is .725 and .805 respectively. The A2 and B1 Gitsaki Collocation Tests (Time 1 & 2), A2 and B1 Progress Collocation Test (Time 2) range from .781 to .864. The test items on the A2 and B1 Gitsaki Collocation Tests were split according to their BBI type and BNC frequency of occurrence (Appendix E, Table A.1 and A.2).

Table 7.3. Reliability coefficients of the research instruments

Test name	Spearman-Brown split-half reliability coefficient	
	Time 1	Time 2
Serial Nonword Recall (A2)	.918	.905
Serial Nonword Recall (B1)	.897	.923
Serial Nonword Recognition (A2)	.831	.812
Serial Nonword Recognition (B1)	.704	.792
L1 Vocabulary Test (A2)	.876	----
L1 Vocabulary Test (B1)	.864	----
A2 Vocabulary Levels Test	.876	.845
B1 Vocabulary Levels Test	.951	.912
A2 Gitsaki Collocation Test	.781	.801
B1 Gitsaki Collocation Test	.852	.864
A2 Progress Vocabulary Test	----	.725
B1 Progress Vocabulary Test	----	.805
A2 Progress Collocation Test	----	.809
B1 Progress Collocation Test	----	.827

Note: Time 1 – January 2008; Time 2 – July 2008.

With regard to construct validity, a good construct should have a theoretical basis that is translated through clear operational definitions of measurable concepts. The constructs that the above-listed tests are meant to tap are presented in detail in Chapter 1, Section 1.2. The constructs in question are well grounded in the existing literature on PSTM, vocabulary and collocations.

Owing to the fact that Serial Nonword Recall and Serial Nonword Recognition correlate at the .01 level, we can argue that the two instruments have high convergent validity. An instrument is thought to have high convergent validity when it correlates strongly with other variables with which it should theoretically correlate. Recall- and recognition-based PSTM measures have been shown to correlate strongly in the existing literature (e.g. Gathercole et al., 1999a). As discussed in Section 7.5, our PSTM measures do not correlate with L1 Vocabulary Test scores, which is in line with a number of existing studies involving adults (e.g. Papagno & Vallar, 1995; Speciale et al., 2004: Experiment 1). This indicates that our PSTM measures have high discriminant validity, i.e. they do not correlate significantly with variables from which they would be expected to differ. The Vocabulary Levels Test was shown to have good construct validity by Laufer & Nation (1999) and Schmitt et al. (2001). The Progress Vocabulary and Collocation Tests have high construct validity in that they reflect vocabulary and collocations introduced and recycled during the TCD language course.

7.5. Correlations among main variables

One of the two aspects the current section focuses on is the strength of correlations among our independent variables⁶⁵. The use of many statistical methods, such as e.g. multiple regression, is typically preceded by the generation and analysis of a correlation matrix, as this can indicate whether or not there are very strong correlations between independent variables to be entered into a multiple regression model. A strong correlation (.80 or stronger) between two or more independent variables in a regression model is referred to as multicollinearity. When independent variables are collinear, the individual significance of each of the independent variables is difficult to assess, regression coefficients become unstable and the size of R is limited (Field, 2005). Having uncorrelated (or weakly correlated) independent variables is, therefore, beneficial.

Apart from dealing with the strength of correlations among the independent variables, this section also looks at the strength of correlations between the independent and the dependent variables. Questions 1, 2, 4 and 5 are addressed in the light of the

⁶⁵ Our independent variables include Serial Nonword Recall scores (Time 1), the Serial Nonword Recognition scores (Time 1), the Exposure scores (between Time 1 and Time 2) and the L1 Vocabulary Test scores (Time 1).

obtained patterns of correlations. It should be pointed out at this stage that the interpretation of correlation coefficients needs to be proceeded with cautiously, as correlation coefficients give no indication of the *direction* of causality. Even though it may be intuitively appealing to conclude that one variable causes the other one to change, in correlational designs there is no *statistical* reason why the direction of causality cannot be opposite to the one that is more intuitively appealing. Therefore, the current section draws no direct conclusions about causality.

Pearson's product-moment formula was used to calculate correlations among the variables that are normally distributed (Section 7.3), while Spearman's formula was used for those variables that have significant deviations from normality. Correlations among the independent variables and vocabulary measures in the A2 and B1 groups are presented first. What follows are matrices with correlations among the independent variables and collocation measures in the A2 and B1 groups. Below each correlation coefficient is presented its significance value. Relevant correlation coefficients are flagged as significant at the .05 level, the .01 level and the .001 level with one, two and three stars respectively.

Correlations among independent variables and L2 vocabulary scores in the A2 group

Four independent variables were identified for the A2 group data, namely the Serial Nonword Recall scores at Time 1, Serial Nonword Recognition at Time 1, the Exposure scores between Time 1 and Time 2 (referred to henceforth as the Exposure scores), and the L1 Vocabulary Test scores at Time 1 (Variables 1, 2, 3 and 4 in Table 7.4 respectively)⁶⁶. The Serial Nonword Recall scores and the Serial Nonword Recognition scores are correlated at the .001 level ($r = .683$), which corresponds to the results obtained in our pilot study (see Chapter 5). The correlation between Serial Nonword Recall (Time 1) and the Exposure scores is non-significant ($r = .281$, $p > .05$), as is the correlation between Serial Nonword Recall (Time 1) and the L1 Vocabulary Test scores (Time 1) ($r = .258$, $p > .05$). The Serial Nonword Recognition scores (Time 1) correlate neither with the Exposure scores ($r = .294$, $p > .05$) nor with the L1

⁶⁶ Variables 1 – 7 in Table 7.4 are normally distributed, and therefore a parametric statistic (Pearson's product-moment correlation coefficient) is used to report correlations among them. Variable 8 is not normally distributed, and therefore a non-parametric statistic (Spearman's correlation coefficient) is used to report correlations between Variable 8 and the remaining variables.

Vocabulary Test scores at Time 1 ($r = .004, p > .05$). The lack of very strong correlations (i.e. $r > .8$) between our independent variables indicates that the variables are not collinear.

Table 7.4. Simple intercorrelations among independent variables and A2 vocabulary measures (n=30)

Variables		1	2	3	4	5	6	7	8
1. Serial Recall (1)		—							
	Sig.								
2. Serial Recognition (1)		.683***	—						
	Sig.	.000							
3. Exposure ^A		.281	.294	—					
	Sig.	.133	.115						
4. L1 Vocabulary Test (1)		.258	.004	.000	—				
	Sig.	.168	.984	.998					
5. A2 Vocabulary Levels Test gains		.459*	.384*	.461*	-.019	—			
	Sig.	.011	.036	.010	.921				
6. A2 Vocabulary Levels Test (1)		.420*	.359	.023	.383*	-.056	—		
	Sig.	.021	.052	.904	.036	.769			
7. A2 Vocabulary Levels Test (2)		.548**	.466*	.164	.366*	.254	.952***	—	
	Sig.	.002	.010	.385	.047	.176	.000		
8. A2 Progress Vocabulary Test (2) ^B		.594**	.528**	.258	.182	.326	.458*	.541**	—
	Sig.	.001	.003	.168	.335	.079	.011	.002	

Note: (1) – scores at Time 1; (2) – scores at Time 2.
^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.
^B The A2 Progress Vocabulary Test scores are non-normally distributed (Spearman's rho).
 * $p < .05$; ** $p < .01$; *** $p < .001$.

The lack of significant correlations between the L1 Vocabulary Test scores (Time 1) and the Serial Nonword Recall / Recognition scores (Time 1) corroborates the finding of Speciale et al. (2004: Experiment 1) that there is no significant link between a Nonword Repetition task⁶⁷ and an L1 (English) low frequency vocabulary knowledge task in adult subjects in an experimental setting ($r = .15, p > .05, n = 38$). The lack of significant correlations between the above-mentioned variables is also in line with the study by Papagno and Vallar (1995), in which Nonword Repetition scores⁶⁸ did not correlate with L1 (Italian) low frequency vocabulary knowledge scores in adults ($r = -.11, p > .05, n = 20$). The measure that both Speciale et al. (2004: Experiment 1) and

⁶⁷ The nonwords selected for inclusion in our PSTM tasks were similar to those used by Speciale et al. (2004, Experiment 1) in the sense that they were of relatively low wordlikeness (please note that the term wordlikeness is used to denote the degree to which nonwords resemble real words in a given language). The Speciale et al. experimental nonwords were constructed by randomly combining a number of consonants and vowels into CV syllables with a view to reflecting neither the subjects' L1 (English) nor L2 (Spanish). In the current study the nonwords for the Serial Nonword Recall and Recognition tasks were created from existing L1 (Polish) CV and CCV syllables, but only the nonwords that obtained the lowest wordlikeness ratings were included in our PSTM measures (see Chapter 5, Section 5.3.1).
⁶⁸ Very little information was provided about the nonwords created for the purpose of this experiment. As a result, we cannot make inferences as to the extent to which the degree of wordlikeness of the Papagno and Vallar nonwords relates to the degree of wordlikeness of our nonwords.

Papagno and Vallar (1995) used to tap PSTM capacity was a Nonword Repetition task. This study extends their finding to two other types of PSTM measures, namely Serial Nonword Recall and Serial Nonword Recognition.

The fact that individual differences in the Serial Nonword Recall and Recognition scores are not related to the scores on the L1 (low frequency) Vocabulary Test scores in the current study echoes the trend presented in the literature that the higher the level of familiarity with L1 the weaker the link between PSTM and L1 vocabulary knowledge becomes. As mentioned in Chapter 3, a number of studies on PSTM and L1 vocabulary knowledge (Gathercole et al., 1992; Papagno & Vallar, 1995; Speciale et al., 2004: Experiment 1) and a study on PSTM and productive use of L2 vocabulary (O'Brien et al., 2006) indicate that the relationship tends to weaken as L1/L2 knowledge increases. Since adults are more proficient L1 users than children, the link between an L1-based Nonword Repetition task and L1 vocabulary knowledge observed in children (e.g. Masoura & Gathercole, 1999) is no longer statistically significant in adult subjects. More advanced language learners are thought to become less reliant on PSTM, as they grow more adept at using their mental lexicons to access phonological representations of close neighbours of a new word to be learnt (Gathercole, 2006a).

The dependent variables include the A2 Vocabulary Levels Test gain scores, the A2 Vocabulary Levels Test scores (Time 1), the A2 Vocabulary Levels Test scores (Time 2) and the A2 Progress Vocabulary Test scores (Time 2) (Variables 5, 6, 7 and 8 respectively in Table 7.4 above). As expected, the A2 Vocabulary Levels Test scores (Time 1) correlate strongly with the A2 Vocabulary Levels Test scores (Time 2) ($r = .952, p < .001$). There is a weak positive correlation between the A2 Vocabulary Levels Test scores (Time 2) and the A2 Vocabulary Levels Test gain scores ($r = .254, p > .05$), but the correlation is non-significant. The A2 Vocabulary Levels Test scores (Time 1) do not correlate with the A2 Vocabulary Levels Test gain scores ($r = -.056, p > .05$). The A2 Progress Vocabulary Test scores correlate with the A2 Vocabulary Levels Test scores (Time 1) and with the A2 Vocabulary Levels Test scores (Time 2) ($r = .458, p < .05$ and $r = .541, p < .01$ respectively; Spearman's rho for nonparametric data).

Both the Serial Nonword Recall and Recognition scores (Time 1) correlate highly with the A2 Vocabulary Levels Test scores (Time 2) ($r = .548, p < .01$ and $r =$

.466, $p < .05$ respectively). The Serial Nonword Recall scores (Time 1) correlate with the A2 Vocabulary Levels Test scores (Time 1) at the .05 level ($r = .420$). The correlation between the Serial Nonword Recognition scores (Time 1) and the A2 Vocabulary Levels Test scores (Time 1) is marginally non-significant ($r = .359$, $p = .052$). A pattern emerges in which the correlation between the two PSTM measures (Time 1) and the A2 Vocabulary Levels Test scores (Time 2) is stronger than that between the two PSTM measures (Time 1) and the A2 Vocabulary Levels Test scores (Time 1). This reflects the pattern of correlations presented in a longitudinal classroom-based study by O'Brien et al. (2006), who used an L1 (English) Serial Nonword Recognition task to tap PSTM capacity in a group of low and high ability⁶⁹ adult learners ($n = 23$ and $n = 20$ respectively) enrolled on an L2 (Spanish) language course. The O'Brien et al. data show that the link between Serial Nonword Recognition at Time 1 and L2 vocabulary (i.e. free productive use of L2 vocabulary in oral production) at Time 1 in the low ability group was significantly weaker ($r = .296$, $p > .05$) than the link between Serial Nonword Recognition at Time 1 and L2 vocabulary scores at Time 2 ($r = .423$, $p < .05$). Our data reveal that the above pattern of correlations detected by O'Brien et al. in their low ability English learners of Spanish can also be observed in A2 Polish learners of English, resident in an English-speaking country, when receptive and controlled productive knowledge of high-frequency L2 vocabulary is measured over a period of 6 months.

The fact that both of our PSTM measures are significantly linked with the L2 Vocabulary Levels Test at Time 2 may indicate that knowledge of high-frequency L2 vocabulary (Time 2) is constrained more by the phonological store than by the articulatory rehearsal component⁷⁰. It should also be pointed out that a comparison of the strength of the correlation between Serial Nonword Recall and high-frequency L2 vocabulary (Time 2) and that between Serial Nonword Recognition and high-frequency L2 vocabulary (Time 2) reveals that the articulatory rehearsal component is more strongly linked to adult L2 vocabulary learning than it is to child L1 vocabulary learning. In a longitudinal study by Gathercole et al. (1999b) involving a group of 4-year-old children learning their L1 the correlation between Serial Nonword

⁶⁹ The exact L2 proficiency level was not specified by the authors.

⁷⁰ Within the WM model a Serial Nonword Recognition task is believed to mainly involve the phonological store, whereas a Serial Nonword Recall task is assumed to involve both the store and the articulatory rehearsal component (Baddeley, 2003).

Recognition (Time 1) and L1 vocabulary knowledge (Time 2) was significantly stronger than that between Nonword Repetition (Time 1) and L1 vocabulary knowledge (Time 2) ($r = .72, p < .001$ and $r = .54, p < .05$ respectively). A reversed pattern observed in our adult L2 learners may be partially caused by the fact that adults approach PSTM memory tasks in a different manner from that of young children. Pickering (2006: 245) reports that while completing PSTM memory tasks children aged four listen passively to memory stimuli without trying to engage in rehearsal. Adults, on the other hand, actively engage in different forms of rehearsal to maintain to-be-remembered information in STM. All of our subjects admitted to rehearsing experimental nonwords while partaking in the Serial Nonword Recall and Recognition tasks. The fact that adults are more aware of the merits of rehearsal, as compared to young children, may also indicate that they are more likely to engage in rehearsal while attempting to put new L2 words to memory. The involvement of the articulatory rehearsal component in L2 vocabulary learning is further explored in Section 7.6.

The Serial Nonword Recall scores (Time 1) and the Serial Nonword Recognition scores (Time 1) correlate with the A2 Vocabulary Levels Test gain scores at the .05 level ($r = .459$ and $r = .384$ respectively). The Serial Nonword Recall scores (Time 1) and the Serial Nonword Recognition scores (Time 1) also correlate with the A2 Progress Vocabulary Test scores (Time 2) ($r = .594$ and $r = .528$ respectively, both $ps < .01$; Spearman's rho for nonparametric data). The strong correlations between our PSTM measures and the A2 Vocabulary Levels Test gain scores link our subjects' Serial Nonword Recall and Recognition scores with the increase in the knowledge of high-frequency L2 vocabulary (i.e. the first and second 1,000 most frequent English words) over a period of six months. The strong correlation between our PSTM measures and the A2 Progress Vocabulary Test scores link the variation in PSTM scores at the commencement of an English language course with the variation in the scores on the end-of-course vocabulary test.

As for the relationship between PSTM and adult L2 vocabulary gains in other longitudinal studies, O'Brien et al. (2006) identified a weak positive correlation between their L1-based Nonword Serial Recognition measure and L2 vocabulary gains, but the correlation was not significant ($r = .341, p > .05, n = 43$). This finding by O'Brien et al. is somewhat surprising (Service, personal communication) considering

the strength of correlations obtained in experimental studies of adults (Gupta, 2003; Papagno & Vallar, 1995). One possible reason why O'Brien et al. did not find a significant correlation between their Serial Nonword Recognition measure and L2 vocabulary gains may have had to do with the nature of their vocabulary measure. O'Brien et al. set out to measure free productive use of L2 vocabulary in oral production, while we focused on testing receptive and controlled productive knowledge of high-frequency L2 vocabulary. In the light of their data O'Brien et al. suggest that PSTM "is not related to the development of productive vocabulary [use] but, rather, to the development of vocabulary knowledge" (O'Brien et al., 2006: 395)⁷¹. The current study confirms the latter part of the above statement, namely that the link between Serial Nonword Recognition (Time 1) and gains in high frequency L2 vocabulary is statistically significant in adult A2 learners of English. The study also shows that there is a significant link between Serial Nonword Recall (Time 1) and L2 vocabulary gains in adults.

The vast majority of the lexical items tested in the A2 Vocabulary Levels Test at Time 1 and Time 2 were not taught explicitly during the TCD English language course the subjects were enrolled on. Considering the fact that the subjects were not enrolled on any other English language course at the time, we suspected that the gains on the A2 Vocabulary Levels Test between Time 1 and Time 2, apart from being linked to the students' PSTM scores, could also be linked to their exposure to highly frequent words outside the classroom. Our data appears to confirm this assumption. The Exposure scores are correlated with the A2 Vocabulary Levels Test gain scores at the .05 level ($r = .461$). The fact that the scores of Exposure to L2 outside the classroom are so closely related to the A2 Vocabulary Levels Test gain scores corroborate the finding by e.g. Durrant (2008) that in experimental conditions the gains of adult subjects embarking on the task of learning L2 word combinations are sensitive to the amount of exposure they receive.

The L1 Vocabulary Test scores (Time 1) do not correlate with the A2 Vocabulary Levels Test gain scores ($r = -.019$, $p > .05$). This is in line with the finding by Papagno and Vallar (1995) that L1 vocabulary knowledge does not correlate with

⁷¹ O'Brien et al. (2006), in the light of their data, hypothesise that PSTM is not linked to the rate of improvement in free productive knowledge L2 vocabulary in spoken output. They, however, do not provide an explanation for this assumption. The current study has not been designed to deal with productive use L2 vocabulary in spoken output.

experimental L2 vocabulary learning in adults ($r = -.11$, $p > .05$). Speciale et al. (2004), mentioned earlier in this section, also included a measurement of L1 vocabulary knowledge in their experimental study exploring the role of PSTM in adult L2 vocabulary learning. However, Speciale et al. did not present correlation coefficients among L1 (English) vocabulary knowledge scores and two sets of L2 (German) vocabulary scores, and therefore it is not clear how our findings relate to theirs. Even though there is no significant link between the L1 Vocabulary Test scores (Time 1) and the A2 Vocabulary Levels Test gains, the L1 Vocabulary Test scores (Time 1) correlate with the A2 Vocabulary Levels Test scores at Time 1 and Time 2 ($r = .383$ and $r = .366$ respectively, $ps < .05$). A strong correlation between an L1 vocabulary measure and an L2 vocabulary measure ($r = .66$, $p < .01$) was reported in a cross-sectional study by Masoura and Gathercole (1999) in children aged 8 to 11 ($n = 45$). The correlation obtained by Masoura and Gathercole in a group of children was significantly stronger than that identified in our adult subjects. Also, in the Papagno and Vallar study (1995) adult polyglot subjects outperformed adult non-polyglot subjects on an L1 low frequency vocabulary task ($\mu = 75.5$ and $\mu = 70.3$ respectively). The difference between L1 vocabulary group means in the Papagno and Vallar study was not significant, but showed a trend towards significance ($p = .07$). As for classroom based-studies on adults, two recently published longitudinal studies on PSTM and adult L2 language learning, i.e. O'Brien et al. (2006) and Speciale et al. (2004, Experiment 2), and a cross-sectional study on PSTM and adult L2 language learning, i.e. Kormos & Sáfár (2008), did not include a measure of L1 vocabulary knowledge, and therefore we cannot compare our finding to their results.

Correlations among independent variables and L2 vocabulary scores in the B1 group

Four independent variables were identified in the B1 group, namely the Serial Nonword Recall scores at Time 1, the Serial Nonword Recognition scores at Time 1, the Exposure scores between Time 1 and Time 2, and the L1 Vocabulary Test scores at Time 1 (Variables 1 – 4 in Table 7.5 below)⁷². The L1 Vocabulary Test does not

⁷² Variables 1 – 7 in Table 7.5 are normally distributed, and therefore a parametric statistic (Pearson's product-moment correlation coefficient) is used to report correlations among them. Variable 8 is not normally distributed, and therefore a non-parametric statistic (Spearman's correlation coefficient) is used to report correlations between Variable 8 and the remaining variables.

correlate with other independent variables (all $ps < .05$). Exposure does not correlate with Serial Nonword Recognition (Time 1), but correlates significantly with Serial Nonword Recall (Time 1) ($r = .555, p < .01$). R of this magnitude does not violate the assumption of absence of multicollinearity among independent variables (Field, 2005). However, a high correlation between two independent variables increases the probability that, when one of them is entered into a multiple regression model with the other one already in it, the variable entered second will be found non-significant and rejected from the model despite possibly being a good predictor of the outcome. The strength of correlation between Serial Nonword Recall (Time 1) and Serial Nonword Recognition (Time 1), i.e $r = .645, p < .001$, is comparable to that obtained in the A2 group.

Table 7.5. Simple intercorrelations among independent variables and B1 vocabulary measures (n=30)

Variables		1	2	3	4	5	6	7	8
1. Serial Recall (1)		—							
	Sig.								
2. Serial Recognition (1)		.645***	—						
	Sig.	.000							
3. Exposure ^A		.555**	.185	—					
	Sig.	.001	.328						
4. L1 Vocabulary Test (1)		-.031	-.006	.018	—				
	Sig.	.871	.973	.926					
5. B1 Vocabulary Levels Test gains (2-1)		.367*	.218	.473**	0.103	—			
	Sig.	.046	.246	.008	0.587				
6. B1 Vocabulary Levels Test (1)		.227	.189	.214	-.024	-.067	—		
	Sig.	.228	.317	.257	.901	.726			
7. B1 Vocabulary Levels Test (2)		.413*	.304	.447*	.021	.036	.908***	—	
	Sig.	.023	.102	.013	.911	.850	.000		
8. B1 Progress Vocabulary Test (2) ^B		.365*	.141	.521**	-.124	.087	.533**	.597***	—
	Sig.	.047	.459	.003	.515	.647	.002	.000	

Note: (1) – scores at Time 1; (2) – scores at Time 2; (2-1) – scores at Time 1 subtracted from scores at Time 2.

^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.

^B The B1 Progress Vocabulary Test scores are non-normally distributed (Spearman's rho).

* $p < .05$; ** $p < .01$; *** $p < .001$.

The lack of a correlation between the L1 Vocabulary Test (Time 1) and Serial Nonword Recall / Recognition (Time 1) replicates the findings obtained in the A2 group. This result corroborates the trend documented in the literature, namely that the greater the familiarity with L1 vocabulary, the weaker its link with PSTM measures becomes (Masoura & Gathercole, 1999; Papagno & Vallar, 1995). As for the relationship between Serial Nonword Recall (Time 1) and Exposure to L2 outside the classroom, the correlation between the two variables is significant at the .01 level. This means that the B1 students with higher Serial Nonword Recall scores (Time 1) received

a significantly higher amount of exposure to L2 outside the classroom as compared to the B1 students with lower Serial Nonword Recall scores (Time 1). This complication could not have been prevented from arising, as our experimental design did not allow us to control the amount of English our subjects were exposed to outside the classroom between Time 1 and Time 2.

Four dependent variables related to L2 vocabulary learning identified in this study are the B1 Vocabulary Levels Test gain scores, the B1 Vocabulary Levels Test scores (Time 1), the B1 Vocabulary Levels Test scores (Time 2) and the B1 Progress Vocabulary Test scores (Time 2) (Variables 5, 6, 7 and 8 respectively in Table 7.5 above). As expected, the B1 Vocabulary Levels Test (Time 1) correlates strongly with the B1 Vocabulary Levels Test (Time 2) ($r = .908, p < .001$). The B1 Vocabulary Levels Test gain scores correlate with the B1 Vocabulary Levels Test scores (Time 2) at the .05 level ($r = .447$), but do not correlate with the B1 Vocabulary Levels Test (Time 1) ($r = -.067, p > .05$). The B1 Progress Vocabulary Test (Time 2) correlates highly with the B1 Vocabulary Levels Test (Time 1) and the B1 Vocabulary Levels Test (Time 2) ($r = .533, p < .01$ and $r = .597, p < .001$; Spearman's rho for nonparametric data).

The correlations between Serial Nonword Recall (Time 1) and L2 vocabulary measures in the B1 group are weaker than those obtained in the A2 group. In the B1 group Serial Nonword Recall (Time 1) is linked with the B1 Vocabulary Levels Test at the .05 level (Time 2) ($r = .413, p = .023$). In the A2 group Serial Nonword Recall (Time 1) and the A2 Vocabulary Levels Test (Time 2) are correlated at the .01 level ($p = .002$). In the B1 group the correlation between Serial Nonword Recall (Time 1) and the B1 Vocabulary Levels gains is significant at the .05 level ($r = .367, p = .046$). In the A2 group the correlation between the corresponding variables is also significant at the .05 level but noticeably stronger ($p = .011$). Unlike the patterns of results obtained in the A2 group, Serial Nonword Recall (Time 1) in the B1 group is not significantly correlated with the B1 Vocabulary Levels Test (Time 1) ($r = .227, p > .05$). In the B1 group the correlation between Serial Nonword Recall (Time 1) and the B1 Progress Vocabulary Test (Time 2) is marginally significant ($r = .365, p = .047$; Spearman's rho for nonparametric data). In the A2 group the A2 Progress Vocabulary Test correlates strongly with Serial Nonword Recall (Time 1) ($p < .01$, Spearman's rho).

The above-mentioned patterns of correlations in our adult subjects resemble the finding by Gathercole et al. (1992) in children ($n = 80$) that the link between L1-based (English) Nonword Repetition and L1 vocabulary knowledge decreases as children's L1 proficiency increases. A comparison of the results obtained by children tested at the age of six ($r = .439, p < .001$) with those recorded two years later ($r = .284, p < .05$) reveals a significant difference in the strength of the correlations. Our data show that a similar pattern can be observed in adult L2 learners as their L2 proficiency increases⁷³.

It should also be noted that although the correlations between Serial Nonword Recall (Time 1) and L2 vocabulary measures (Time 2 and gain scores) in the B1 group are significant at the .05 level, they become non-significant when Exposure is partialled out (all $ps > .05$). This is caused by the fact that there is a high correlation between Serial Nonword Recall (Time 1) and Exposure in the B1 group. Since both Serial Nonword Recall (Time 1) and Exposure correlate with L2 vocabulary measures (Time 2 and gain scores), it is statistically impossible to separate the influence of Serial Nonword Recall (Time 1) and Exposure on L2 vocabulary scores.

Our second PSTM measure, i.e. the Serial Nonword Recognition measure (Time 1), does not correlate with the B1 Vocabulary Levels Test gains ($r = .218, p = .246$), the B1 Vocabulary Levels Test (Time 1) ($r = .189, p = .317$) or the B1 Vocabulary Levels Test at Time 2 ($r = .304, p = .102$), and the B1 Vocabulary Progress Test ($r = .141, p > .05$; Spearman's rho). This corresponds to the results obtained by O'Brien et al. (2006) in their high ability⁷⁴ group of adult L2 learners ($n = 20$). The L1-based (English) Serial Nonword Recognition task employed by O'Brien et al. failed to correlate with a measure of free productive use of L2 (Spanish) vocabulary in oral production at Time 1 ($r = .411, p > .05$) and Time 2 ($r = .175, p > .05$). Their Serial Nonword Recognition task also failed to correlate with gain scores in free productive use of L2 vocabulary ($r = -.050, p > .05$).

⁷³ The Gathercole et al. (1992) followed a group of 80 children of the same age over the period of four years. Since the timeframe of our project did not allow us to re-test our A2 learners of English when they have progressed to the B1 level, we opted for testing two separate groups of adult subjects, one at the A2 level of L2 proficiency and the other at the B1 level. As mentioned in Section 7.3, the two groups of subjects did not differ significantly in their PSTM scores, their L1 low frequency vocabulary scores or their Exposure scores.

⁷⁴ The exact L2 proficiency level was not specified.

The comparison of the simple correlations between the Serial Nonword Recall task (Time 1) and L2 vocabulary measures and the simple correlations between the Serial Nonword Recognition task (Time 1) and L2 vocabulary measures indicates that the two PSTM measures yield different results at the B1 level. The two PSTM tasks are highly correlated in both A2 and B1 groups, and yet at the B1 level Serial Nonword Recall is significantly linked with L2 vocabulary measures, while Serial Nonword Recognition is not. Assuming that Serial Nonword Recall taps both the articulatory component and the nonarticulatory storage component of the PSTM, while Serial Nonword Recognition taps only the latter (Baddeley, 2003a: 197; O'Brien et al., 2006: 395; Snowling et al., 1991), our findings may indicate that the articulatory component of the PSTM remains involved in L2 vocabulary learning in adult subjects at the B1 level of L2 proficiency, while the nonarticulatory storage component may no longer be involved. However, the correlations between Serial Nonword Recall (Time 1) and L2 vocabulary measures (Time 2 and gain scores) become non-significant when Exposure is partialled out (all p s > .05). This may mean that neither the nonarticulatory storage component nor the articulatory storage component is involved in L2 vocabulary learning at the B1 level of proficiency. This issue is further explored in Section 7.6.

As in the A2 group, the Exposure variable in the B1 group is significantly linked to the B1 Vocabulary Levels Test gain scores ($r = .473$, $p < .01$) and the B1 Vocabulary Levels Test scores at Time 2 ($r = .447$, $p < .05$). The above correlations indicate that the B1 learners with a higher level of exposure to L2 outside the classroom (between Time 1 and Time 2) obtained significantly higher scores on the 2,000 and 3,000 Vocabulary Levels Test at Time 2, and that their gains on the test were also significantly higher than the gains of the remaining B1 students. All of our subjects obtained higher scores on the 2,000 section of the Vocabulary Levels Test as compared to the scores obtained on the 3,000 section of the Vocabulary Levels Test, which lends support to Meara's (1992) frequency model of vocabulary learning (see Chapter 3, Section 3.3.3) within the specified frequency bands. The Exposure variable in the B1 group also correlates strongly with the B1 Progress Vocabulary Test (Time 2) ($r = .521$, $p < .01$; Spearman's rho).

The last of our independent variables, i.e. the L1 Vocabulary Test designed to measure productive knowledge of L1 low frequency words, does not correlate with any

of our dependent variables in the B1 group. As has been mentioned, in the A2 group the L1 Vocabulary Test (Time 1) correlates with the A2 Vocabulary Levels Test at Time 1 and Time 2 at the .05 level. The fact that the link between the L1 Vocabulary Test and Vocabulary Levels Test (Time 1 and 2) is significant in the A2 group but not in the B1 group may indicate that individual differences in knowledge of L1 vocabulary are more of a factor in individual differences in knowledge of L2 high frequency vocabulary at the beginning stage of L2 learning than at later stages.

Correlations among independent variables and L2 collocation scores in the A2 group

Four independent variables referred to in this section are Serial Nonword Recall (Time 1), Serial Nonword Recognition (Time 1), Exposure to L2 outside the classroom (between Time 1 and Time 2), and the L1 Vocabulary Test scores (Time 1). The variables in question are listed in Table 7.6 below as Variables 1, 2, 3 and 4 respectively⁷⁵. The dependent variables include A2 Gitsaki Collocation Test gain scores, A2 Gitsaki Collocation Test scores (Time 1), A2 Gitsaki Collocation Test scores (Time 2) and A2 Progress Collocation Test scores (Time 2) (Variables 5, 6, 7 and 8 respectively in Table 7.6 below). The A2 Gitsaki Collocation Test scores (Time 1) correlate highly with the A2 Gitsaki Collocation Test scores (Time 2) ($r = .932$, $p < .001$). The strength of this correlation resembles that obtained between the A2 Vocabulary Levels Test scores (Time 1) and the A2 Vocabulary Levels Test scores (Time 2). This implies that subjects who obtained higher vocabulary and collocation scores at Time 1 tended to obtain higher scores at Time 2 as well. There are also significant correlations between the A2 Progress Collocation Test (Time 2) and the A2 Gitsaki Collocation Test (Time 1 and Time 2) ($r = .440$ and $r = .427$ respectively, both $ps < .05$; Spearman's rho).

⁷⁵ Since correlations among the independent variables in the A2 group have already been presented, they are not discussed here (see the section entitled "Correlations among independent variables and L2 vocabulary scores in the A2 group").

Table 7.6. Simple intercorrelations among independent variables and A2 collocation measures (n=30)

Variables	1	2	3	4	5	6	7	8
1. Serial Recall (1)	—							
Sig.								
2. Serial Recognition (1)	.683***	—						
Sig.	.000							
3. Exposure ^A	.281	.294	—					
Sig.	.133	.115						
4. L1 Vocabulary Test (1)	.258	.004	.000	—				
Sig.	.168	.984	.998					
5. A2 Gitsaki Collocation Test gains (2-1)	.494**	.436*	.292	.130	—			
Sig.	.006	.016	.118	.492				
6. A2 Gitsaki Collocation Test (1)	.375*	.281	.095	.290	.132	—		
Sig.	.041	.132	.619	.119	.488			
7. A2 Gitsaki Collocation Test (2)	.512**	.408*	.190	.304	.482**	.932**	—	
Sig.	.004	.025	.314	.102	.007	.000		
8. A2 Progress Collocation Test (2) ^B	.362*	.370*	.051	-.078	.079	.440*	.427*	—
Sig.	.049	.044	.788	.681	.675	.015	.019	

Note: (1) – scores at Time 1; (2) – scores at Time 2; (2-1) – scores at Time 1 subtracted from scores at Time 2.

^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.

^B The A2 Progress Collocation Test scores are non-normally distributed (Spearman's rho).

*p < .05; **p < .01; ***p < .001.

Serial Nonword Recall (Time 1) correlates with the A2 Gitsaki Collocation Test (Time 1) ($r = .375$, $p < .05$), the A2 Gitsaki Collocation Test (Time 2) ($r = .512$, $p < .01$) and the A2 Progress Collocation Test (Time 2) ($r = .362$, $p = .049$; Spearman's rho). Serial Nonword Recognition (Time 1) correlates significantly with the A2 Gitsaki Collocation Test (Time 2) ($r = .408$, $p < .05$) and the A2 Progress Collocation Test (Time 2) ($r = .370$, $p = .044$; Spearman's rho). There is a weak positive correlation between Serial Nonword Recognition (Time 1) and the A2 Gitsaki Collocation Test (Time 1), but it is non-significant ($r = .281$, $p > .05$). The correlation between both PSTM measures (Time 1) and A2 Gitsaki Collocation Test (Time 2) is stronger than that between the PSTM measures (Time 1) and the A2 Gitsaki Collocation Test (Time 1). The strength of the link between our Serial Nonword Recognition measure (Time 1) and the A2 Gitsaki Collocation Test (Time 1 and 2) resembles the strength of the correlations between a measure of the use of subordinate clauses and an L1-based Serial Nonword Recognition task used by O'Brien et al. (2006) in a longitudinal, classroom-based study of L2 learning in adult learners. In a group of low ability students ($n = 23$) the strength of the correlation between L1-based Serial Nonword Recognition at Time 1 and use of L2 subordinate clauses at Time 1 and Time 2 was $r = .381$ ($p > .05$) and $r = .471$ ($p < .05$) respectively.

Both Serial Nonword Recall (Time 1) and Serial Nonword Recognition (Time 1) correlate highly with the A2 Gitsaki Collocation Test gain scores. However, the link between Serial Nonword Recall (Time 1) and the A2 Gitsaki Collocation Test gain scores is stronger than that between Serial Nonword Recognition (Time 1) and the A2 Gitsaki Collocation Test gain scores ($r = .494$, $p < .01$ and $r = .436$, $p < .05$). The fact that both Serial Nonword Recognition (tapping the phonological store) and Serial Nonword Recall (tapping the phonological store and the articulatory component) correlate with the A2 Gitsaki Collocation Test gain scores could be interpreted as an indication that gains in knowledge of L2 collocations are linked more closely to the functioning of the phonological store than to the articulatory component. It should also be noted that in this study the measure indexing both the phonological store and the articulatory component, i.e. Serial Nonword Recall, is a better predictor of the subsequent gains in L2 collocational knowledge over the period of 6 months than the measure indexing the phonological store only, i.e. Serial Nonword Recognition. The issue of the strength of the relationship between the learning of L2 collocations and the two components of PSTM is further investigated in Section 7.7.

Neither Exposure nor the L1 Vocabulary Test correlates significantly with any of the collocation variables. Exposure was shown to correlate with the A2 Vocabulary Levels Test gain scores at the .05 level earlier in this section. The correlation between Exposure and the A2 Gitsaki Collocation Test gain scores is also positive, but it is not significant ($r = .292$, $p = .118$). The lack of significant correlation between the two variables may be brought about by the nature of our Exposure measure. The Exposure measure reflects an average number of hours per day of exposure to L2 outside the classroom. The number of hours was recorded by students on a daily basis during selected six weeks between Time 1 and Time 2. The measure, therefore, involved a subjective estimation of the amount of exposure to L2 outside the classroom at the duration of the TCD language course.

To assess the impact of exposure to L2 collocations on L2 collocational knowledge an additional analysis was carried out, in which our subjects' knowledge of selected high-frequency and low-frequency collocations on the A2 Gitsaki Test (Time 2) was compared. The collocations in the A2 Gitsaki Collocation tests were ranked according to their BNC frequencies (16 – 16,090, see Table A.1 in Appendix E). 15

collocations with the highest BNC (British National Corpus) scores and 15 collocations with the lowest BNC scores were selected to create two sets of scores, one representing high-frequency collocations and the other low-frequency collocations. The collocations were also ranked in terms of their MI3 (Mutual Information cubed) scores (8.267 – 24.782). 15 collocations with the highest MI3 scores and 15 collocations with the lowest MI3 scores were selected to create two sets of scores, one representing strong collocations and the other weak collocations. The A2 subjects' knowledge of high-frequency collocations at Time 2 was significantly higher than their knowledge of low-frequency collocations ($p < .01$), and the subjects' knowledge of strong collocations was also significantly higher ($p < .05$). This extends Meara's (1992) frequency model of vocabulary learning (see Chapter 3, Section 3.3.3) to L2 collocation knowledge.

Correlations among independent variables and L2 collocation scores in the B1 group

The independent variables referred to in this section are Serial Nonword Recall (Time 1), Serial Nonword Recognition (Time 1), Exposure to L2 outside the classroom, and the L1 Vocabulary Test (Time 1)⁷⁶. The dependent variables include the following: the B1 Gitsaki Collocation Test gains, the B1 Gitsaki Collocation Test (Time 1), the B1 Gitsaki Collocation Test (Time 2) and the B1 Progress Collocation Test (Time 2). B1 Gitsaki Collocation Test (Time 1) correlates with B1 Gitsaki Collocation Test (Time 2) ($r = .975$, $p < .001$; see Table 7.7 below), which resembles the strength of correlation between the A2 Gitsaki Collocation Test scores at Time 1 and Time 2. The B1 Gitsaki Collocation Test gain scores correlate neither with the B1 Gitsaki Collocation Test scores (Time 1) nor with the B1 Gitsaki Collocation Test scores (Time 2). The L1 Vocabulary Test and the Exposure measure fail to correlate with any of our L2 collocation measures (all $ps > .05$).

⁷⁶ Since correlations among the independent variables in the B1 group have already been presented, they are not discussed here (see the section entitled "Correlations among independent variables and L2 vocabulary scores in the B1 group").

Table 7.7. Simple intercorrelations among independent variables and B1 collocation measures (n=30)

Variables		1	2	3	4	5	6	7	8
1. Serial Recall (1)		—							
	Sig.								
2. Serial Recognition (1)		.645***	—						
	Sig.	.000							
3. Exposure ^A		.555**	.185	—					
	Sig.	.001	.328						
4. L1 Vocabulary Test (1)		-.031	-.006	.018	—				
	Sig.	.871	.973	.926					
5. B1 Gitsaki Collocation Test gains (2-1)		.460*	.296	.184	-.118	—			
	Sig.	.011	.113	.330	.534				
6. B1 Gitsaki Collocation Test (1)		.312	.287	.186	-.098	.044	—		
	Sig.	.093	.124	.326	.606	.816			
7. B1 Gitsaki Collocation Test (2)		.403*	.343	.220	-.121	.264	.975***	—	
	Sig.	.027	.064	.243	.525	.159	.000		
8. B1 Progress Collocation Test (2) ^B		.439*	.199	.306	-.169	.221	.314	.459*	—
	Sig.	.015	.289	.100	.372	.241	.091	.011	

Note: (1) – scores at Time 1; (2) – scores at Time 2; (2-1) – scores at Time 1 subtracted from scores at Time 2.

^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.

^B The B1 Progress Collocation Test scores are non-normally distributed (Spearman's rho).

* $p < .05$; ** $p < .01$; *** $p < .001$.

Serial Nonword Recall (Time 1) correlates with the B1 Gitsaki Collocation Test gain scores ($r = .460$, $p < .05$), the B1 Gitsaki Collocation Test scores (Time 2) ($r = .403$, $p < .05$) and the B1 Progress Collocation Test (Time 2) ($r = .439$, $p < .05$; Spearman's rho). The correlation between Serial Nonword Recall (Time 1) and the B1 Gitsaki Collocation Test scores (Time 1) is non-significant ($r = .312$, $p > .05$). Serial Nonword Recognition (Time 1) does not correlate with the B1 Gitsaki Collocation Test scores (Time 1 and Time 2), the B1 Gitsaki Collocation Test gain scores and the B1 Progress Collocation Test scores (all $ps > .05$). In the O'Brien et al. (2006) study of adult L2 learning a significant correlation was reported between L1-based Serial Nonword Recognition (Time 1) and L2 subordinate clause use (Time 2) in a high ability group of learners ($r = .532$, $p < .05$). In the current study the correlation between Serial Nonword Recognition (Time 1) and the B1 Gitsaki Collocation Test (Time 2) shows a trend towards significance ($p = .064$), while the correlation between Serial Nonword Recall (Time 1) and the B1 Gitsaki Collocation Test (Time 2) is significant at the .05 level.

A comparison of patterns of correlations obtained in the A2 and B1 groups reveals that the strength of the correlations between our PSTM measures and L2 collocation measures also decreases as L2 proficiency increases. In the A2 group Serial

Nonword Recall correlates with the A2 Gitsaki Collocation Test (Time 2) and with the A2 Gitsaki Collocation Test gains at the .01 level, and Serial Nonword Recognition correlates with the above collocation measures at the .05 level. In the B1 group Serial Nonword Recall correlates with the B1 Gitsaki Collocation Test (Time 2), the B1 Gitsaki Collocation Test gains, and the B1 Progress Collocation Test the .05 level (even when Exposure is partialled out). When the strength of the correlations between Serial Nonword Recognition and L2 collocation measures is compared in the A2 group and the B1 group, weakening also becomes apparent. However, it has to be noted that the correlations between Serial Nonword Recognition and L2 collocation measures in the B1 group are all non-significant.

The fact that, when Exposure is partialled out, Serial Nonword Recall (Time 1) in the B1 group no longer correlates with the B1 Vocabulary Levels Test gain scores and B1 Progress Collocation Test scores, but it remains significantly linked with the B1 Gitsaki Collocation Test gain scores / B1 Progress Collocation Test scores (both $ps < .05$) could be interpreted as support for O'Brien et al.'s (2006: 399) claim that at post-beginning stages of L2 learning PSTM, instead of being deployed for L2 vocabulary learning, is deployed for learning more complex structures. L2 vocabulary learning at post-elementary levels of L2 proficiency can be aided by existing L2 lexical knowledge in semantic, conceptual and phonological dimensions, and therefore there is less need for PSTM involvement (Gathercole, 2006a: 515). In the case of L2 collocation learning at post-elementary levels, a greater learning effort can be expected as compared to the learning of single words (Verstraten, 1992: 28). Consequently, even though the link between the Serial Nonword Recall task and L2 vocabulary learning diminishes at the B1 level, the link between Serial Nonword Recall and L2 collocation learning remains strong.

The fact that Serial Nonword Recall (Time 1), but not Serial Nonword Recognition (Time 1), correlates with L2 collocation learning can be interpreted as evidence that the phonological store component of PSTM ceases to be involved in L2 collocation learning at the B1 level, while the articulatory rehearsal component remains involved. In an experimental study by Martin (2009: 90) a measure of L2 word order correlated with a recall-based PSTM measure but failed to correlate with a recognition-based PSTM measure in adult learners at the A1 level of L2 proficiency. This study

demonstrates that the same pattern of correlations can be observed between L2 collocation learning and recall-/recognition-based PSTM measures in B1 adult learners of English. The issue of the strength of the relationship between L2 collocation learning and the articulatory rehearsal component of PSTM is further investigated in Section 7.7.

7.6. Regression models of the link between PSTM and L2 vocabulary

This section presents multiple regression models representing the link between PSTM (tapped by L1-based Serial Nonword Recall and Recognition tasks) and L2 vocabulary learning (see Questions 1 and 2 in Section 7.2). A curve estimation function in SPSS 16.0 was used to check linearity of the relationships in our data before our each of our regression models was calculated. The multiple regression models presented below include up to three independent variables. Including a higher number of variables would have resulted in overfitting our regression models. Overfitting is said to occur when there are too many independents in relation to sample size. It should also be pointed out that the sample size we used in our analyses (i.e. $n=30$ in each proficiency group) allows the identification of large effects only (Field, 2005).

The correlation coefficient squared (R^2) is presented in this section as an estimation of the amount of variance in the dependent variables that is accounted for by the independent variables within each regression model. As much as R^2 is a very useful measure of the importance of an effect, it cannot be used to infer *causal* relationships between variables (Field, 2005: 129). As a result, this section makes no direct conclusions about causality.

PSTM and L2 vocabulary scores in the A2 group

When the Serial Nonword Recall scores (Time 1) were entered into a simple linear regression model as an independent variable and the A2 Vocabulary Levels Test gain scores as a dependent variable, the model explained 21 per cent of the variance in the A2 Vocabulary Levels Test gain scores ($R^2 = .210$). An adjusted R^2 of .182 indicates that if the model were derived from the population from which the sample was taken, the model would explain 18 per cent of the variance in the A2 Vocabulary Levels Test gain scores. When the Serial Nonword Recognition scores (Time 1) were

entered into a simple linear regression model as an independent variable, the model accounted for 15 per cent of the variance in the A2 Vocabulary Levels Test gain scores ($R^2 = .147$, adj. $R^2 = .117$). Our data shows that in adult A2 learners of English, resident in an English-speaking country (Ireland), Serial Nonword Recall and Serial Nonword Recognition (Time 1) account for 21 and 15 per cent of variance in the 1,000 and 2,000 Vocabulary Levels Test⁷⁷ gain scores over a period of six months respectively.

A number of hierarchical multiple regressions were carried out with the A2 Vocabulary Levels Test gain scores as a dependent variable and the Exposure scores, the L1 Vocabulary Test scores and Serial Nonword Recall / Recognition scores (Time 1) as independent variables. Since most of the A2 Levels Test items were not explicitly taught in class, scores reflecting the amount of exposure to L2 outside the classroom (average number of hours per day between Time 1 and 2) were included as an independent variable in hierarchical multiple regressions described below. The Exposure scores and the L1 Vocabulary Test scores were added in Block 1 and Block 2 respectively. The Serial Nonword Recall scores (Time 1) were added in Block 3 in the first hierarchical multiple regression, while the Serial Nonword Recognition scores were added in Block 3 in the second hierarchical multiple regression (see Table 7.8 below). When the Exposure scores and the L1 Vocabulary Test scores were included in the model, the model accounted for 21 per cent of the variance in the A2 Vocabulary Levels Test gains and it was significant at the .05 level ($F(2,27) = 3.658$, $p < .05$). A comparison of R^2 and adjusted R^2 values indicates shrinkage of 5.8 per cent ($R^2 = .213$, adj. $R^2 = .155$). When Serial Nonword Recall (Time 1) was added in Block 3, R^2 and adjusted R^2 values rose to .34 and .27 respectively and the model remained significant at the .05 level ($F(3,26) = 4.544$). When Serial Nonword Recognition (Time 1) was added in Block 3, R^2 and adjusted R^2 values reached .28 and .20 respectively, and the model was also significant at the .05 level ($F(3,26) = 3.380$).

⁷⁷ The relevant sections of the Vocabulary Test probe knowledge of English words from the 1,000 and 2,000 frequency bands.

Table 7.8. Summary of hierarchical multiple regressions for variables predicting gains in A2 Vocabulary Levels Test gain scores (n=30)

Variables	R	R ²	Adj. R ²	df	F	Sig.
A2 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.461*	.212	.185	1, 28	7.571	.010
L1 Vocabulary Test (1)	.462*	.213	.155	2, 27	3.658	.039
Serial Recall (1)	.586*	.344	.268	3, 26	4.544	.011
A2 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.461*	.213	.185	1, 28	7.571	.010
L1 Vocabulary Test (1)	.462*	.213	.155	2, 27	3.658	.039
Serial Recognition (1)	.530*	.281	.198	3, 26	3.380	.033
A2 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.461*	.213	.185	1, 28	7.571	.010
Serial Recall (1)	.575**	.331	.281	2, 27	6.669	.004
A2 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.461*	.213	.185	1, 28	7.571	.010
Serial Recognition (1)	.529*	.280	.227	2, 27	5.255	.012
A2 Vocabulary Levels Test gains (2-1)						
Serial Recognition (1)	.384*	.147	.117	1, 28	4.836	.036
Serial Recall (1)	.469*	.219	.162	2, 27	3.807	.035
A2 Vocabulary Levels Test gains (2-1)						
Serial Recall (1)	.459*	.211	.182	1, 28	7.473	.011
Serial Recognition (1)	.469*	.219	.162	2, 27	3.807	.035

Note: (1) – scores at Time 1; (2) – scores at Time 2; (2-1) – scores at Time 1 subtracted from scores at Time 2.

^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.

*p < .05; **p < .01; ***p < .001.

As can be seen in Table 7.8 above, the L1 Vocabulary Test contributed only minimally to the increase in R² (R² difference = .01) after Exposure had been added to the model. When Exposure and Serial Nonword Recall (Time 1) were added in Block 1 and Block 2 respectively (and the L1 Vocabulary Test was excluded), the model became significant at the .01 level (*F*(2,27) = 6.669) and it explained 33 per cent of the variance of the A2 Vocabulary Levels Test gain scores (R² = .331, adj. R² = .281). Serial Nonword Recall (Time 1) explained additional 11.8 per cent of variance in the A2 Vocabulary Levels Test gain scores. When Exposure was added in Block 1 and Serial Nonword Recognition (Time 1) in Block 2 (and the L1 Vocabulary Test was excluded), the model remained significant at the .05 level (*F*(2,27) = 5.255) and it explained 28 per cent of the variance of the A2 Vocabulary Levels gain scores (R² = .280, adj. R² = .227). Serial Nonword Recognition accounted for additional 6.7 per cent of variance in the A2 Vocabulary Levels Test gain scores.

As discussed in Section 7.6, the A2 Vocabulary Levels Test gain scores correlate significantly stronger with Serial Nonword Recall (Time 1) than with Serial Nonword Recognition (Time 1). Both PSTM tasks are believed to tap the phonological store, but only Serial Nonword Recall is thought to index the articulatory rehearsal component. In order to investigate the contribution of the articulatory rehearsal

component to predicting the A2 Vocabulary Levels Test gain scores, we carried out two hierarchical multiple regressions (presented in Table 7.8 above). When Serial Nonword Recognition (Time 1) was entered in Block 1 and Serial Nonword Recall (Time 1) was entered in Block 2, Serial Nonword Recall (Time 1) explained additional 7.2 per cent of variance in the dependent variable (R^2 difference = .072, adj. R^2 difference = .045). When the entry order was reversed, Serial Nonword Recognition (Time 1), entered in Block 2, explained additional 0.8 per cent of variance (R^2 difference = .008, adj. R^2 difference = -.020). Considering the fact that in the current study the amount of variance in the A2 Vocabulary Levels Test gain scores accounted for by Serial Nonword Recognition (the phonological store) is 14.7 per cent, we could argue that in adult L2 learning at the A2 level of proficiency both the phonological store and the rehearsal component are actively involved, which is in keeping with the Baddeley and Hitch WM model.

The amount of additional variance that Serial Nonword Recall (Time 1) accounted for after Serial Nonword Recognition (Time 1) had been added to the model (i.e. 7.2 per cent) can be attributed to the articulatory rehearsal component. The fact that the amount of additional variance explained by Serial Nonword Recall (Time 1) is considerably higher than the amount of additional variance explained by Serial Nonword Recognition (Time 1) – 7.2 per cent and 0.8 per cent respectively – can be interpreted as an indication that the articulatory rehearsal component contributes to gains in the knowledge of L2 vocabulary in adult learners. As mentioned earlier, in the case of young children the link between L1 vocabulary and a recognition-based PSTM measure has been shown to be stronger than that between L1 vocabulary and a recall-based PSTM measure (Gathercole et al., 1999b). In an experimental study by Martin (2009: 87) of adult artificial language learning, on the other hand, the strength of the correlation of the productive knowledge of L2 adjectives and prepositions with a recall-based PSTM measure ($r = .352$ and $r = .334$ respectively, $p < .05$) was stronger than that with a recognition-based PSTM measure ($r = .261$ and $r = .283$ respectively, $ps > .05$). This difference between young children and adults can be related to the finding that adults, unlike young children (Gathercole & Hitch, 1993; Pickering, 2006: 245), make use of rehearsal strategies in order to hold novel items in memory (Pressley & Van Meter, 1992). Considering the fact that in the current study the amount of variance in the A2 Vocabulary Levels Test gain scores accounted for by the phonological store

(14.7 per cent) is significantly higher than the amount of additional variance accounted for by the articulatory rehearsal component (7.2 per cent), we could assume that the phonological store is more closely linked to adult L2 vocabulary learning, but that the articulatory rehearsal component plays a role as well.

In relation to the contribution of the phonological store and the articulatory rehearsal component to the A2 Vocabulary Levels Test scores (Time 2), the hierarchical multiple regressions presented in Table 7.9 above indicate that when Serial Nonword Recognition (Time 1) and Serial Nonword Recall (Time 1) were entered in Block 1 and Block 2 respectively, Serial Nonword Recall (Time 1) accounted for an additional 9.9 per cent of the variance in the dependent variable. When Serial Nonword Recall (Time 1) was added in Block 1 instead, Serial Nonword Recognition (Time 1) explained an additional 1.5 per cent of the variance. This very much resembles the pattern of results described in the preceding paragraph.

In order to assess how much variance in the A2 Vocabulary Levels Test scores (Time 2) the PSTM measures explain, a number of hierarchical multiple regressions with the A2 Vocabulary Levels Test scores (Time 2) as a dependent variable were carried out (see Table 7.9 below). When Exposure was added in Block 1 and the L1 Vocabulary Test in Block 2, the L1 Vocabulary Test explained an additional 13 per cent of the variance in the dependent variable. However, the model was not significant ($F(2,27) = 2.588$, $p > .05$). When Serial Nonword Recall was added in Block 3, the model became significant at the .01 level ($F(3,26) = 4.779$) and it explained 35 per cent of variance in the dependent variable ($R^2 = .335$, adj. $R^2 = .281$). The Serial Recall accounted for additional 19 per cent of variance within that model. When Serial Nonword Recognition was added in Block 3, the model was also significant at the .01 level ($F(3,26) = 4.669$).

Table 7.9. Summary of hierarchical multiple regressions for variables predicting the A2 Vocabulary Levels Test scores at Time 2 (n=30)

Variables	R	R ²	Adj. R ²	df	F	Sig.
A2 Vocabulary Levels Test scores (2)						
Exposure ^A	.164	.027	-.008	1, 28	0.778	.385
L1 Vocabulary Test (1)	.401	.161	.099	2, 27	2.588	.094
Serial Recall (1)	.596**	.355	.281	3, 26	4.779	.009
A2 Vocabulary Levels Test scores (2)						
Exposure ^A	.164	.027	-.008	1, 28	0.778	.385
L1 Vocabulary Test (1)	.401	.161	.099	2, 27	2.588	.094
Serial Recognition (1)	.592**	.350	.275	3, 26	4.669	.009
A2 Vocabulary Levels Test scores (2)						
L1 Vocabulary Test (1)	.366*	.134	.103	1, 28	4.323	.047
Serial Recall (1)	.595**	.355	.307	2, 27	7.416	.003
A2 Vocabulary Levels Test scores (2)						
L1 Vocabulary Test (1)	.366*	.134	.103	1, 28	4.323	.047
Serial Recognition (1)	.591**	.349	.301	2, 27	7.246	.003
A2 Vocabulary Levels Test scores (2)						
Serial Recognition (1)	.466*	.217	.189	1, 28	7.751	.010
Serial Recall (1)	.562**	.316	.266	2, 27	6.248	.006
A2 Vocabulary Levels Test scores (2)						
Serial Recall (1)	.548**	.301	.276	1, 28	12.047	.002
Serial Recognition (1)	.562**	.316	.266	2, 27	6.248	.006

Note: (1) – scores at Time 1; (2) – scores at Time 2.

^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.

*p < .05; **p < .01; ***p < .001.

When the L1 Vocabulary Test was added in Block 1 instead of Exposure (Table 7.9 above), the L1 Vocabulary Test was significantly linked with the A2 Vocabulary Levels Test at Time 2 ($F(1,28) = 4.323$, $p = .047$). However, when Serial Nonword Recall (Time 1) was added in Block 2, the b-value of L1 Vocabulary Test became non-significant ($p > .05$), while b-value of Serial Nonword Recall (Time 1) was highly significant ($p < .01$). When Serial Nonword Recognition (Time 1) was added in Block 2 in the second set of hierarchical multiple regressions, both the L1 Vocabulary Test and Serial Nonword Recognition made a significant contribution to predicting the A2 Vocabulary Levels Test scores at Time 2 ($p < .05$ and $p < .01$ respectively). In a study by Masoura and Gathercole (1999) the link between L1 vocabulary and L2 vocabulary was found to be highly significant in children aged 8 to 11. The b-value of the L1 Vocabulary Test in the former regression model of ours was non-significant, which may indicate that in adult subjects the link between L1 vocabulary and L2 vocabulary is significantly weaker as compared to that observed in children.

PSTM and L2 vocabulary scores in the B1 group

When Serial Nonword Recall scores (Time 1) were entered into a simple linear regression model as an independent variable and the B1 Vocabulary Levels Test gain scores as a dependent variable, the model was significant ($F(1,28) = 4.364$, $p = .046$) and it explained 13 per cent of the variance in the A2 Vocabulary Levels Test gain scores ($R^2 = .135$, adj. $R^2 = .104$). When Serial Nonword Recognition scores (Time 1) were entered into a simple linear regression model as an independent variable, the model accounted for 5 per cent of the variance in the B1 Vocabulary Levels Test gain scores ($R^2 = .048$, adj. $R^2 = .014$), and it was not significant ($F(1,28) = 1.404$, $p > .05$).

To assess the degree to which the two remaining independent variables (i.e. the Exposure scores and the L1 Vocabulary Test scores) account for the variance in the dependent variable (i.e. the B1 Vocabulary Levels Test gain scores), two hierarchical multiple regressions were carried out (see Table 7.10 below). When Exposure and the L1 Vocabulary Test (Time 1) were added in Block 1 and in Block 2 respectively, the L1 Vocabulary Test explained less than 1 per cent of variance in the B1 Vocabulary Levels Test gain scores. The b-value of the L1 Vocabulary Test was not significant, and the variable was excluded from subsequent multiple regressions. When Exposure was added in Block 1 and Serial Nonword Recall (Time 1) was added in Block 2, the model was significant at the .05 level ($F(2,27) = 4.244$). When Serial Nonword Recognition (Time 1) was added in Block 2 instead of Serial Nonword Recall (Time 1), the model was also significant at the .05 level ($F(2,27) = 4.288$). Despite the fact that the above two multiple regression models were significant at the .05 level, the b-value of Serial Nonword Recall (Time 1) and the b-value of Serial Nonword Recognition (Time 1) turned out to be non-significant within the models ($p = .459$ and $p = .433$ respectively). This implies that within the regression models neither of the two PSTM tasks (Time 1) was a good predictor of the outcome. The b-value of Serial Nonword Recall was significant within a simple linear regression model, but became non-significant within a multiple regression model. As anticipated in Section 7.5, a high correlation between Serial Nonword Recall (Time 1) and Exposure resulted in Serial Nonword Recall (Time 1) being rejected from a multiple regression model when Exposure had already been added to it. Serial Nonword Recall (Time 1) may be a good predictor of the B1 Vocabulary Levels Test gain scores, but its high correlation with

Exposure measure makes it impossible to disentangle the impact that these two independent variables may have on the gains on the B1 Vocabulary Levels Test between Time 1 and Time 2.

Table 7.10. Summary of hierarchical multiple regressions for variables predicting gains in B1 Vocabulary Levels Test scores (n=30)

Variables	R	R ²	Adj. R ²	df	F	Sig.
B1 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.473**	.223	.196	1, 28	8.049	.008
L1 Vocabulary Test (1)	.482*	.232	.175	2, 27	4.085	.028
Serial Recall (1)	.499	.249	.163	3, 26	2.880	.055
B1 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.473**	.223	.196	1, 28	8.049	.008
L1 Vocabulary Test (1)	.482*	.232	.175	2, 27	4.085	.028
Serial Recognition (1)	.500	.250	.164	3, 26	2.894	.054
B1 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.473**	.223	.196	1, 28	8.049	.008
Serial Recall (1)	.489*	.239	.183	2, 27	4.244	.025
Serial Recognition (1)	.494	.244	.156	3, 26	2.791	.060
B1 Vocabulary Levels Test gains (2-1)						
Exposure ^A	.473**	.223	.196	1, 28	8.049	.008
Serial Recognition (1)	.491*	.241	.185	2, 27	4.288	.024
Serial Recall (1)	.494	.244	.156	3, 26	2.791	.060

Note: (1) – scores at Time 1; (2) – scores at Time 2; (2-1) – scores at Time 1 subtracted from scores at Time 2.

^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.

*p < .05; **p < .01; ***p < .001.

In order to assess the amount of variance that our independent variables explain in the B1 Vocabulary Levels Test scores (Time 2), two additional hierarchical multiple regressions carried out (see Table 7.11 below). When Exposure was added in Block 1 and Serial Nonword Recall (Time 1) in Block 2, the model was significant at the .05 level ($F(2,27) = 4.235$) and accounted for 23.9 per cent of variance in the dependent variable ($R^2 = .239$, adj. $R^2 = .182$). Serial Nonword Recall (Time 1) explained additional 4 per cent of the variance. When Serial Nonword Recognition (Time 1) was added in Block 2 instead of Serial Nonword Recall (Time 1), the model was also significant at the .05 level ($F(2,27) = 4.506$) and it explained 25 per cent of variance in the dependent variable ($R^2 = .250$, adj. $R^2 = .195$). Serial Nonword Recognition (Time 1) accounted for additional 4.9 per cent of variance in the B1 Vocabulary Levels Test scores (Time 2) within that model. However, the b-values of Serial Nonword Recall and Serial Nonword Recognition were non-significant within both models ($p = .248$ and $p = .187$ respectively). When Serial Nonword Recall (Time 1) was entered into a simple linear regression as an independent variable, the b-value of Serial Nonword Recall was significant at the .05 level and the model accounted for 17 per cent of the variance in the B1 Vocabulary Levels Test scores (Time 2). The fact that the value of

Serial Nonword Recall was significant in a simple regression and became non-significant in a multiple regression can again be linked to its strong correlation with the Exposure measure.

Table 7.11. Summary of hierarchical multiple regressions for variables predicting the B1 Vocabulary Levels Test scores at Time 2 (n=30)

Variables	R	R ²	Adj. R ²	df	F	Sig.
B1 Vocabulary Levels Test scores (2)						
Exposure ^A	.447*	.199	.171	1, 28	6.975	.013
Serial Recall (1)	.489*	.239	.182	2, 27	4.235	.025
B1 Vocabulary Levels Test scores (2)						
Exposure ^A	.447*	.199	.171	1, 28	6.975	.013
Serial Recognition (1)	.500*	.250	.195	2, 27	4.506	.020

Note: (1) – scores at Time 1; (2) – scores at Time 2.
^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.
 *p < .05; **p < .01; ***p < .001.

As discussed earlier, one other longitudinal classroom-based project apart from ours (O'Brien et al., 2006) examined the link between PSTM and L2 vocabulary gains in post-beginner adults. O'Brien et al. discovered no significant link in their high ability group⁷⁸ between an L1-based Serial Nonword Recognition task and gains on a measure of free productive use of L2 words in spoken output ($r = -.050$, $p > .05$). Our simple linear regressions show that in adult B1 learners of English, resident in an English-speaking country (Ireland), Serial Nonword Recall (Time 1) accounts for 13 per cent of variance in the 2,000 and 3,000 Vocabulary Levels Test gain scores and for 17 per cent of the variance in the 2,000 and 3,000 Vocabulary Levels Test scores (Time 2). However, Serial Nonword Recognition (Time 1), in line with the results obtained by O'Brien et al. in relation to free productive use of L2 vocabulary, is not significantly linked to either the 2,000 and 3,000 Vocabulary Levels Test gain scores or the 2,000 and 3,000 Vocabulary Levels Test scores (Time 2).

However, when Serial Nonword Recall (Time 1) and Exposure were entered into multiple regressions with either the B1 Vocabulary Levels Test gain scores or the B1 Vocabulary Levels Test scores (Time 2) as a dependent variable, the b-values of both Serial Nonword Recall (Time 1) and Exposure become non-significant (all $ps > .05$) irrespective of the entry order. This indicates that when the amount of exposure to L2 outside the classroom is partialled out, Serial Nonword Recall (Time 1) is no longer

⁷⁸ Please note that the exact proficiency level was not specified.

a good predictor of the B1 Vocabulary Levels Test gain scores or the B1 Vocabulary Levels Test scores (Time 2).

The fact that the link between Serial Nonword Recall (Time 1) and L2 vocabulary measures (Time 2) becomes non-significant when Exposure is partialled out, corroborates the trend reported in the literature that the strength of the relationship between recall-based PSTM measures and vocabulary learning diminishes as language proficiency increases. Our results resemble the pattern of results obtained by Cheung (1996) in relation to L2 vocabulary learning (tapped via a paired-associate learning task) among children with bigger L2 vocabularies. It seems reasonable to interpret this finding as an indication that at the B1 level of proficiency neither low capacity of the phonological store nor low efficiency of the rehearsal process significantly impairs the learning of L2 vocabulary.

7.7. Regression models of the link between PSTM and L2 collocations

This section contains simple linear regression models and multiple regression models representing the link between PSTM (tapped by our Serial Nonword Recall and Recognition tasks) and L2 collocation learning (see Questions 4 and 5 in Section 7.2). The section focuses on the data from the A2 group first, and then moves on to the data obtained in the B1 group.

PSTM scores and L2 collocation scores in the A2 group

When the Serial Nonword Recall scores (Time 1) were entered into a simple linear regression model as an independent variable, the model accounted for 24.4 per cent of the variance in the A2 Gitsaki Collocation Test gain scores ($R^2 = .244$, adj. $R^2 = .217$), and it was significant at the .01 level ($F(1,28) = 9.015$). When the Serial Nonword Recognition scores (Time 1) were entered into a simple linear regression model as an independent variable and the A2 Gitsaki Collocation Test gain scores as a dependent variable, the model was significant at the .05 level ($F(1,28) = 6.565$) and it explained 18.9 per cent of the variance in the A2 Gitsaki Collocation Test gain scores ($R^2 = .189$, adj. $R^2 = .161$). Since there was a weak positive relationship between Exposure and the A2 Gitsaki Collocation Test gains (Section 7.5), Exposure was entered into hierarchical multiple regressions (see Table 7.12 below) in order to see

how much variance in the dependent variable our PSTM measures and Exposure measure would account for together. When Exposure was entered in Block 1 and Serial Nonword Recall in Block 2, the model explained 26.9 per cent of variance in the dependent variable. When Serial Nonword Recognition was entered in Block 2 instead, the model explained 21.9 per cent of variance. Both models were significant at the .05 level, but the b-value of Exposure within the models was, as expected, non-significant.

Table 7.12. Summary of hierarchical multiple regressions for variables predicting gains in A2 Gitsaki Collocation Test gain scores (n=30)

Variables	R	R ²	Adj. R ²	df	F	Sig.
A2 Gitsaki Collocation Test gains (2-1)						
Exposure ^A	.292	.085	.052	1, 28	2.602	.118
Serial Recall (1)	.519*	.269	.215	2, 27	4.967	.015
A2 Gitsaki Collocation Test gains (2-1)						
Exposure ^A	.292	.085	.052	1, 28	2.602	.118
Serial Recognition (1)	.468*	.219	.161	2, 27	3.789	.035
A2 Gitsaki Collocation Test gains (2-1)						
Serial Recognition (1)	.436*	.189	.161	1, 28	6.565	.016
Serial Recall (1)	.512*	.262	.207	2, 27	4.789	.017
A2 Gitsaki Collocation Test gains (2-1)						
Serial Recall (1)	.494**	.244	.217	1, 28	9.015	.006
Serial Recognition (1)	.512*	.262	.207	2, 27	4.789	.017

Note: (1) – scores at Time 1; (2) – scores at Time 2; (2-1) – scores at Time 1 subtracted from scores at Time 2.
^A Exposure to L2 outside the classroom was measured on six instances in the period between Time 1 and 2.
 *p < .05; **p < .01; ***p < .001.

As indicated in Section 7.5, the A2 Gitsaki Collocation Test gain scores are more strongly associated with Serial Nonword Recall (Time 1) than with Serial Nonword Recognition (Time 1), which may be explicable in terms of the PSTM components that each of the tasks is believed to index. In order to investigate the contribution of the articulatory rehearsal component to predicting the A2 Gitsaki Collocation Test gain scores, we carried out two hierarchical multiple regressions (see Table 7.12 above). When Serial Nonword Recognition (Time 1) was entered in Block 1 and Serial Nonword Recall (Time 1) was entered in Block 2, Serial Nonword Recall (Time 1) explained additional 7.3 per cent of variance in the dependent variable. When the entry order was reversed, Serial Nonword Recognition (Time 1) entered in Block 2 explained only additional 1.8 per cent of variance. The additional amount of variance that Serial Nonword Recall (Time 1) accounted for after Serial Nonword Recognition (Time 1) had been added to the model can be attributed to the articulatory rehearsal component. The fact that the additional amount of variance explained by Serial Nonword Recall (Time 1) is higher than the additional amount of variance explained by

Serial Nonword Recognition (Time 1) appears to suggest that in adults the articulatory rehearsal component contributes to gains in the knowledge of L2 collocations.

This lends support to Gupta's (2003) suggestion that the strength of the correlations between recall- / recognition-based PSTM tasks and child L1 learning may differ from the strength of the correlations between recall- / recognition-based PSTM tasks and adult L2 learning owing to the differences between these two populations in the use of strategies in recall- and recognition-based PSTM tasks. The fact that the link between the rehearsal component of PSTM and vocabulary learning is observable in adults but not in young children (Gathercole et al., 1999b) may be related to the finding that adults, unlike young children (Gathercole & Hitch, 1993; Pickering, 2006: 245; Pressley & Van Meter, 1992), use rehearsal strategies to hold novel phonological information in memory.

This study shows that L2 collocation learning is more strongly linked with Serial Nonword Recall (Time 1) than it is with Serial Nonword Recognition (Time 1). The fact that there is a stronger correlation between the Serial Nonword Recall measure and the A2 Gitsaki Collocation Test gain scores than between the Serial Nonword Recognition measure and the A2 Gitsaki Collocation Test gain scores is in line with an experimental study by Martin (2009: 90), in which some grammatical measures (e.g. adjective-noun word order) in adult artificial foreign language learning were more strongly linked to Nonword Repetition ($ps < .05$) than to Serial Nonword Recognition ($ps > .05$). The additional amount of variance in the A2 Gitsaki Collocation Test gain scores explained by Serial Nonword Recall (the articulatory rehearsal component) (7.3 per cent) is lower than the variance explained by Serial Nonword Recall (the phonological store) (18.9 per cent). This resembles the pattern of results obtained in relation to the A2 Vocabulary Levels Test gain scores.

As Table 7.13 below shows, L2 collocation knowledge (Time 2) is also more strongly linked to Serial Nonword Recall (Time 1) than it is to Serial Nonword Recognition (Time 1) ($p < .01$ and $p < .05$ respectively). When Serial Nonword Recognition (Time 1) and Serial Nonword Recall (Time 1) were added in Block 1 and Block 2 respectively, Serial Nonword Recall (Time 1) accounted for additional 10.2 per cent of variance in the A2 Gitsaki Collocation Test scores (Time 2). The additional amount of explained variance showed a trend towards significance ($p = .063$). When the

entry order was reversed, Serial Nonword Recognition explained additional 0.7 per cent of variance in the dependent variable. This resembles the pattern observed in relation to the A2 Gitsaki Collocation Test gain scores described above. The additional amount of variance in the A2 Gitsaki Collocation Test scores (Time 2) explained by the articulatory rehearsal component (10.2 per cent, $p > .05$) is again lower than the variance explained by the phonological store (16.7 per cent, $p < .05$).

Table 7.13. Summary of hierarchical multiple regressions for variables predicting A2 Gitsaki Collocation Test scores at Time 2 (n=30)

Variables	R	R ²	Adj. R ²	df	F	Sig.
A2 Gitsaki Collocation Test scores (2)						
L1 Vocabulary Test (1)	.304	.093	.060	1, 28	2.857	.102
Serial Recall (1)	.542**	.294	.241	2, 27	5.616	.009
A2 Gitsaki Collocation Test scores (2)						
L1 Vocabulary Test (1)	.304	.093	.060	1, 28	2.857	.102
Serial Recognition (1)	.508*	.258	.203	2, 27	4.702	.018
A2 Gitsaki Collocation Test scores (2)						
Serial Recognition (1)	.408*	.167	.137	1, 28	5.599	.025
Serial Recall (1)	.518*	.269	.214	2, 27	4.956	.015
A2 Gitsaki Collocation Test scores (2)						
Serial Recall (1)	.512**	.262	.236	1, 28	9.945	.004
Serial Recognition (1)	.518*	.269	.214	2, 27	4.956	.015

Note: (1) – scores at Time 1; (2) – scores at Time 2.
 * $p < .05$; ** $p < .01$; *** $p < .001$.

PSTM scores and L2 collocation scores in the B1 group

When Serial Nonword Recall (Time 1) was entered into a simple linear regression model as an independent variable, the model accounted for 21.1 per cent of the variance in the B1 Gitsaki Collocation Test gain scores ($R^2 = .211$, adj. $R^2 = .183$), and it was significant at the .05 level ($F(1,28) = 7.507$). When Serial Nonword Recognition scores (Time 1) were entered into a simple linear regression model as an independent variable and the B1 Gitsaki Collocation Test gain scores as a dependent variable, the model explained 8.7 per cent of the variance in the A2 Gitsaki Collocation Test gain scores ($R^2 = .087$, adj. $R^2 = .055$), but the amount of variance explained was not significant ($F(1,28) = 2.683$, $p > .05$). This implies the B1 Gitsaki Collocation Test gain scores are more strongly linked with Serial Nonword Recall (Time 1) than they are with Serial Nonword Recognition (Time 1), which may be explicable in terms of the PSTM components that each of the tasks is thought to index within the Working Memory model.

In order to investigate the contribution of the articulatory rehearsal component to predicting the B1 Gitsaki Collocation Test gain scores and to predicting the B1 Gitsaki Collocation Test scores (Time 2), we carried out a number of hierarchical multiple regressions, two of which are presented in Table 7.14 below. When the B1 Gitsaki Collocation Test gain scores were entered into a regression model as a dependent variable, Serial Nonword Recognition (Time 1) was entered in Block 1 and Serial Nonword Recall (Time 1) was entered in Block 2, Serial Nonword Recall (Time 1) explained an additional 12.4 per cent of variance in the dependent variable ($p < .05$). When the order of entry was reversed, Serial Nonword Recognition (Time 1), entered in Block 2, explained less than 0.01 per cent of variance. The additional amount of variance in the dependent variable accounted for by the articulatory rehearsal component (Serial Nonword Recall in Block 2) was significant ($p < .05$). This could be interpreted as an indication that the lower efficiency of the articulatory rehearsal component constrains the learning of L2 collocations in the B1 group more than it does the learning of L2 vocabulary.

Table 7.14. Summary of hierarchical multiple regressions for variables predicting B1 Gitsaki Collocation Test scores at Time 2 and gain scores (n=30)

Variables	R	R ²	Adj. R ²	df	F	Sig.
B1 Gitsaki Collocation Test gain scores						
Serial Recognition (1)	.296	.087	.055	1, 28	2.683	.113
Serial Recall (1)	.460*	.211	.153	2, 27	3.620	.040
B1 Gitsaki Collocation Test scores (2)						
Serial Recognition (1)	.343	.117	.086	1, 28	3.727	.064
Serial Recall (1)	.417	.174	.113	2, 27	2.843	.076

Note: (1) – scores at Time 1; (2) – scores at Time 2.
 * $p < .05$; ** $p < .01$; *** $p < .001$.

7.8. Crossed-lagged correlational analyses of the link between PSTM and L2 vocabulary/collocations

This section addresses the issue of whether it is PSTM that causally affects L2 vocabulary/collocation knowledge or whether the reverse holds true. As discussed in Chapter 2, two alternative accounts propose opposite directions of causality in the above-mentioned relationship. The PSTM account (Baddeley et al., 1998) posits that PSTM causally determines both nonword recall/recognition ability and vocabulary knowledge/learning. Within this account a correlation between nonword recall/recognition ability and vocabulary knowledge/learning is believed to reflect the effect of the underlying variable, i.e. PSTM capacity. The linguistic account (Snowling et al., 1991), on the other hand, proposes that it is vocabulary size that is a causal determinant of nonword recall/recognition ability. Within this account the focus is on factors other than PSTM that influence performance on nonword tasks, namely the extent of lexical/phonological knowledge that an individual possesses. Both accounts acknowledge that the level of lexical/phonological knowledge may affect performance on nonword tasks (see e.g. a discussion of the concepts of wordlikeness and redintegration in Chapter 2). The major difference between the two accounts is that the PSTM account presents PSTM as causally linked with the growth of vocabulary knowledge, while the linguistic account posits that the involvement of PSTM may be relatively minor.

The method used in the current study to investigate the issue of causality between PSTM and L2 vocabulary/collocation knowledge in L2 adult learners is a cross-lagged panel correlation paradigm. By way of background, the rudiments of the method require two constructs, X and Y, measured at two different points in time. As shown in Figure 7.14 below, the two constructs generate four variables (Variable X at Time 1, Variable X at Time 2, Variable Y at Time 1 and Variable Y at Time 2) and six correlations among the four variables. Two of the six correlations, $P_{X_1Y_2}$ and $P_{X_2Y_1}$, are referred to as cross-lagged. $P_{X_1Y_2}$ represents a correlation between Variable X at Time 1 and Variable Y at Time 2, while $P_{X_2Y_1}$ represents a correlation between Variable X at Time 2 and Variable Y at Time 1.

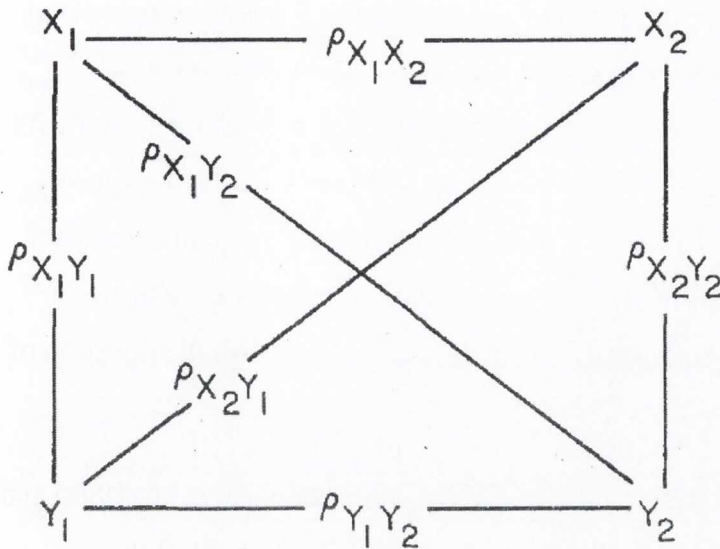


Figure 7.14. Cross-lagged panel correlation paradigm (from Kenny, 1975: 888)

According to the assumptions of a cross-lagged correlational analysis, the correlation should be stronger in the causal than in the noncausal direction (e.g. Campbell, 1988; Campbell & Stanley, 1963; Crano et al., 1972; Crano & Mellon, 1978). To test for the significance of the differences between the cross-lagged correlations in this study a formula attributed to Pearson and Filon (cited in Kenny, 1975: 896) was used. Since the formula is not easily accessible, it is reproduced in Equation 7.1 below:

$$z = \frac{\sqrt{N}(r_{14} - r_{23})}{\sqrt{(1 - r_{14}^2)^2 + (1 - r_{23}^2)^2 - k}}$$

$$k = (r_{12} - r_{24}r_{14})(r_{34} - r_{24}r_{23}) + (r_{13} - r_{12}r_{23})(r_{24} - r_{12}r_{14})$$

$$+ (r_{12} - r_{13}r_{23})(r_{34} - r_{13}r_{14}) + (r_{13} - r_{14}r_{34}) + (r_{24} - r_{34}r_{23})$$

Equation 7.1. Testing the significance of the differences between cross-lagged correlations

N – sample size

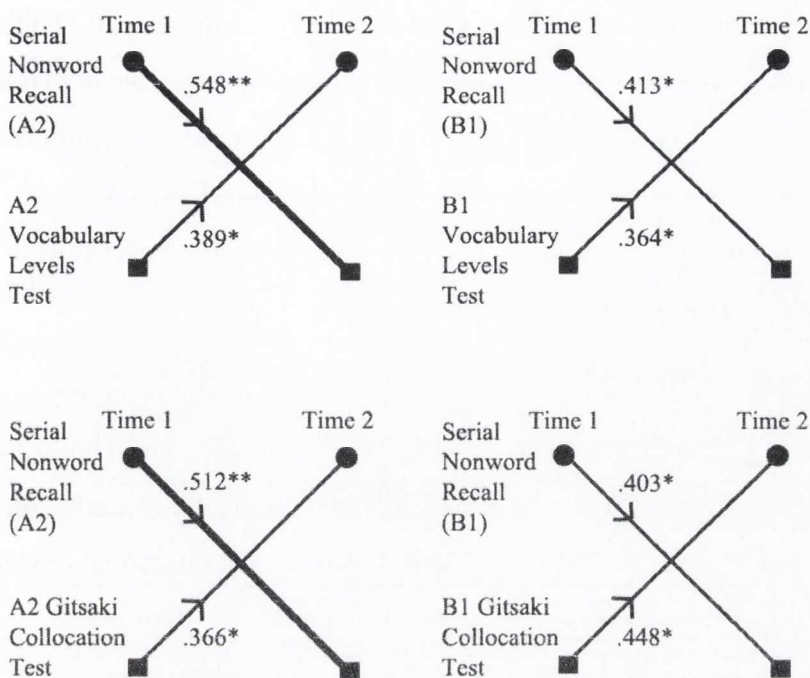
1, 2, 3, 4 – variables x_1, y_1, x_2, y_2

r – correlation coefficient

Cross-lagged correlations between Serial Nonword Recall (Time 1) and L2 vocabulary/collocation scores (Time 2) in the A2 and B1 groups ($r = .403$ to $.548$) indicate that in young adult learners there is a significant link between earlier Serial Nonword Recall performance and later performance on L2 vocabulary/collocation measures. As outlined in Sections 7.5, 7.6 and 7.7, the strength of correlations between

pairs of relevant measures appears to decrease as L2 proficiency increases ($p < .01$ at the A2 level, $p < .05$ at the B1 level). Cross-lagged correlations between L2 vocabulary/collocation scores (Time 1) and Serial Nonword Recall (Time 2) in the A2 and B1 groups ($r = .364$ to $.448$, all $ps < .05$) indicate that there is also a significant relationship between earlier performance on L2 vocabulary/collocation measures and later Serial Nonword Recall performance. These findings are consistent with *any* hypothesis that links the variables in question, and do not yet establish the direction of causality.

In order to discriminate between the PSTM account (Baddeley et al., 1998) and the linguistic account (Snowling et al., 1991), we tested the significance of the differences between the obtained cross-lagged correlations (see Figure 7.15 below). The cross-lagged correlation between Serial Nonword Recall (Time 1) and the A2 Vocabulary Levels Test scores (Time 2) is significant at the .01 level ($r = .548$), and appears stronger than the converse cross-lagged correlation between the A2 Vocabulary Levels Test scores (Time 1) and Serial Nonword Recall (Time 2). The difference between the cross-lagged correlation coefficients in the A2 group is significant ($z = 2.497$, $p < .05$). The cross-lagged correlation between Serial Nonword Recall (Time 1) with the A2 Gitsaki Collocation Test scores (Time 2) ($r = .512$, $p < .05$) is also significantly stronger than the cross-lagged correlation between the A2 Gitsaki Collocation Test scores (Time 1) and Serial Nonword Recall (Time 2) ($z = 2.049$, $p < .05$). The correlations in the B1 group show a different pattern, though. The correlation between Serial Nonword Recall (Time 1) and the B1 Vocabulary Levels Test (Time 2) is not significantly different from the correlation between the B1 Vocabulary Levels Test (Time 1) and Serial Nonword Recall (Time 2) ($r = .413$ and $r = .364$ respectively, both $ps < .05$; $z = -0.198$, $p > .05$). The same applies to cross-lagged correlations between Serial Nonword Recall and B1 Gitsaki Collocation Test scores ($z = -0.680$, $p > .05$).



Note: lines in bold denote a cross-lagged correlation that is significantly higher than the other cross-lagged correlation.

Figure 7.15. Cross-lagged correlations between Serial Nonword Recall and vocabulary/ collocation scores in A2 and B1 groups

The pattern of results obtained at the A2 level can be interpreted as support for the PSTM account, according to which PSTM is a causal determinant of subsequent vocabulary knowledge. The finding by Gathercole et al. (1992) in relation to PSTM and L1 learning in children aged 4 and 5 can, therefore, be extended to early stages of L2 vocabulary and collocation learning in adult L2 learners. Our data also support a recent computational model by Gupta and Tisdale (2009), which demonstrates that variation in PSTM capacity has an impact on the speed of vocabulary learning (Simulation 5). According to the Gupta and Tisdale model “PSTM functionality is a critical and causal determinant of phonological vocabulary learning – without PSTM functionality, there is no vocabulary learning/growth” (Gupta & Tisdale, 2009: 496).

The pattern of results obtained at the B1 level also appears to be in keeping with the PSTM account, according to which the link between PSTM and vocabulary learning is likely to weaken as language proficiency increases. Our data suggest that at the B1 level of proficiency the cross-lagged correlations between Serial Nonword Recall and L2 vocabulary/collocation knowledge are no longer significantly different, which can be interpreted as a weakening of the influence of PSTM. If the correlations

between L2 vocabulary/collocation scores (Time 1) and Serial Nonword Recall (Time 2) were statistically higher than the correlations between Serial Nonword Recall (Time 1) and the L2 vocabulary/collocation scores (Time 2), we could argue – in keeping with the linguistic account – that L2 lexical/phonological knowledge causally determines subsequent Serial Nonword Recall performance and that the involvement of PSTM is minimal. Our data does not offer support for that viewpoint.

The fact that the correlations between the L2 vocabulary/collocation scores (Time 1) and Serial Nonword Recall (Time 2) are significant, however, could be interpreted as an indication of the influence of the lexical/phonological knowledge on subsequent PSTM performance. In the Gupta and Tisdale (2009) computational model, for example, Nonword Repetition ability increases as a function of increasing vocabulary knowledge and increasing knowledge of phonological regularities (Simulation 4). It has to be noted, however, that in this study Serial Nonword Recall is not L2-based, and therefore it does not share the phonological content with L2 vocabulary. As a result, the impact of L2 lexical/phonological knowledge (Time 1) on performance on L1-based Serial Nonword Recall (Time 2) should be expected to be minimal. Even though L2 vocabulary/collocation knowledge (Time 1) is significantly linked with our L1-based Serial Nonword Recall (Time 2), it does not cause an enhanced performance on Serial Nonword Recall task at Time 2. There is no significant increase in our subjects' performance on Serial Nonword Recall between Time 1 and Time 2. This is in keeping with a study by French and O'Brien (2008) which shows that non-L2-based Nonword Repetition task scores in children learning L2 do not change significantly over time, while L2-based Nonword Repetition task scores do. We suspect, therefore, that a significant correlation between L2 vocabulary/collocation knowledge (Time 1) and Serial Nonword Recall (Time 2) arises out of the fact that the Serial Nonword Recall scores (Time 1) and the Serial Nonword Recall scores (Time 2) are very highly correlated ($r = .967$ in the A2 group, $r = .952$ in the B1 group). Both Serial Nonword Recall (Time 1) and Serial Nonword Recall (Time 2) are significantly linked with L2 vocabulary/collocation scores (Time 1). The scores that our subjects obtained on Serial Nonword Recall task remain very stable over time.

The comparison of the strength of cross-lagged correlations between Serial Nonword Recognition (Time 1) and L2 vocabulary/collocation scores (Time 2) and the

converse cross-lagged correlation between L2 vocabulary/collocation scores (Time 1) and Serial Nonword Recognition (Time 2) indicates that there are no significant differences between the pairs of correlations presented in Figure 7.16 below (all p s > .05). In the A2 group Serial Nonword Recognition (Time 1) correlates with both the A2 Vocabulary Levels Test (Time 2) ($r = .466, p < .01$) and with the A2 Gitsaki Collocation Test (Time 2) ($r = .408, p < .05$). In the B1 group Serial Nonword Recognition (Time 1) correlates neither with the B1 Vocabulary Levels Test (Time 2) nor the B1 Gitsaki Collocation Test (Time 2) ($r = .304$ and $r = .343$ respectively, p s > .05). There are weak positive correlations between L2 vocabulary/collocation scores (Time 1) and the Serial Nonword Recognition scores (Time 2), but the correlations are in all cases non-significant (all p s > .05).

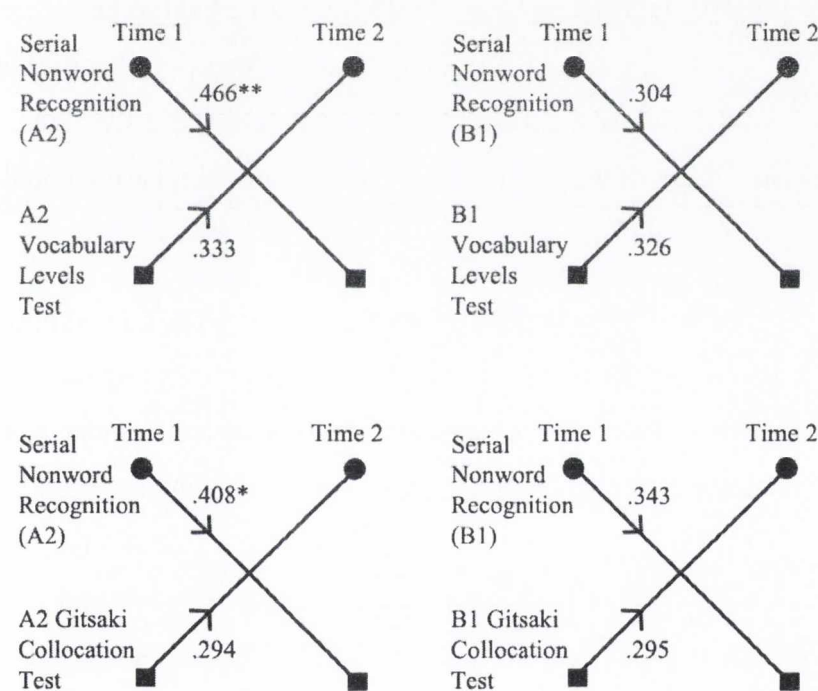


Figure 7.16. Cross-lagged correlations between Serial Nonword Recognition and vocabulary/ collocation scores in A2 and B1 groups

Figure 7.16 above differs noticeably from Figure 7.15 presented earlier in this section in that it does not contain any lines in bold denoting significantly higher cross-lagged correlations. Even though the correlation between Serial Nonword Recognition (Time 1) and the A2 Vocabulary Levels Test (Time 2) appears to be stronger than the correlation between the A2 Vocabulary Levels Test (Time 1) and Serial Nonword Recognition (Time 2) ($r = .466, p < .01$ and $r = .333, p > .05$ respectively), the

difference between the correlations is not significant ($z = 1.254, p > .05$). The same applies to the difference between the pairs of cross-lagged correlations involving Serial Nonword Recall (Time 1 and 2) and the A2 Gitsaki Collocation Test (Time 1 and 2) ($z = 1.008, p > .05$). There are three possible explanations as to why no significant differences between pairs of cross-lagged correlations were discovered at the A2 level in relation to the Serial Nonword Recognition scores, while significant differences were detected in relation to the Serial Nonword Recall scores. First, it is plausible that neither of the two components of PSTM is on its own causally linked to subsequent L2 vocabulary/collocation knowledge, and that only when a PSTM measure that taps *both* components is used does PSTM emerge as a causal determiner of subsequent L2 vocabulary/collocation knowledge. Second, it is also plausible that the correlations involving the Serial Nonword Recognition measure would have turned out to be stronger, had we had access to a larger sample. Our sample size, as mentioned earlier in this chapter, allows the identification of large effects only. Third, the satisfactory but lower than ideally desired reliability of the Serial Nonword Recognition measure could also have a bearing on the interpretation of the results.

The relatively low reliability of the Serial Nonword Recognition task in adults, as compared to the high reliability of the Serial Nonword Recall task, brings up the question of whether Serial Nonword Recognition taps the same construct in adults as it is believed to tap in children. The Serial Nonword Recognition task has mainly been used in research on PSTM and child L1 learning, and has not yet been extensively explored in studies of adult L2 learning. The task was employed by Gathercole and colleagues in a series of innovative studies (e.g. Gathercole et al., 2001) to tap PSTM capacity in children who might have (even minor) articulatory difficulties. To the best of our knowledge, the Serial Nonword Recognition task has so far been used in only one other longitudinal study of adult L2 learning (O'Brien et al., 2006), and the study in question did not aim to compare a Serial Nonword Recognition procedure with a more widely used Serial Nonword Recall procedure. In a very recent study of artificial L2 learning in adults (Martin, 2009) a very low correlation between the Serial Nonword Recognition task and the Nonword Repetition task was detected ($r = .34$), which the author interprets as a possible problem with reference to the construct validity of the Serial Nonword Recognition task (Martin, 2009: 47-48). Despite the fact that in this study the correlation between Serial Nonword Recall and Serial Nonword Recognition

was significant at the .01 level, we feel it is essential that the issue of construct validity of the Serial Nonword Recognition task when used in an adult population be explored in future studies.

Chapter 8: Conclusions

8.1. A summary of the current research findings

The present chapter deals with three issues. First, it provides an overview of how the study dealt with the six research questions which were posed in Chapter 1. Related pairs of these questions are considered in each of three subsections below. For each pair of questions, we begin with a brief account of the relevant research context, and then proceed to an assessment of whether the present study has resolved the particular issues raised or has contributed in some way to the advancement of knowledge on the topic in question. In the following section, the limitations of the study are set out under a number of headings. Finally, Section 3 considers some of the issues which should be pursued in future research.

8.1.1. PSTM capacity and L2 vocabulary/collocation knowledge and learning in A2 and B1 adult L2 learners

Question 1⁷⁹ and Question 4⁸⁰ address the issue of the strength of the relationship between PSTM and knowledge/learning of L2 vocabulary/collocation in adults at the A2 and B1 levels of L2 proficiency.

L2 vocabulary knowledge and learning

It is a well-documented phenomenon that the involvement of PSTM in L1 vocabulary knowledge/learning diminishes with increased familiarity with L1 (Gathercole et al., 1992; Papagno & Vallar, 1995; Speciale et al., 2004: Experiment 1). As mentioned in Section 2.5.1, at the onset of L1 learning, each word encountered by children is a *novel* word form. On the basis of the increasingly robust L1 phonological knowledge that children gradually build up, phonological generalizations start to emerge (Edwards et al., 2004: 434). As this process continues, children attempting to learn a novel L1 word can access phonological representations of close neighbours of

⁷⁹ **Question 1:** Does the relationship between PSTM and L2 vocabulary learning (in class and outside the classroom) decrease with an increased level of L2 proficiency in adults, as has been shown to happen in the case of child L1 learning?

⁸⁰ **Question 4:** In the case of adults, is there a relationship between PSTM and the learning of L2 collocations (in class and outside the classroom) at the A2 and B1 levels of L2 proficiency? If so, does the relationship between PSTM and L2 collocation learning decrease with an increased level of L2 proficiency?

that word, thus lowering the burden placed on limited-capacity PSTM. The Gupta and Tisdale (2009) computational model of STM simulates the process of learning phonological regularities from the input. It shows that when phonological knowledge becomes robust, it contributes to subsequent learning of novel words irrespective of programmed PSTM efficiency (what the authors call PSTM functioning). This paints a picture of PSTM as a language learning device (e.g. Baddeley et al., 1998) that has a strong impact on vocabulary learning/knowledge at the initial stages of L1 learning (e.g. Gathercole et al., 1992).

The mechanism of novel word learning within a particular linguistic system is explored in the Gupta and MacWhinney (1997) model of PSTM, a computational representation of the phonological loop from the Hitch and Baddeley WM model. When a novel word form is presented auditorily, the phonemes making up the word are activated one by one. When recognition processes fail to activate any node at the phonological chunk layer, a new chunk node representing the new word is activated. The syllable template chunks phonemes into syllables and syllables into a word via a number of cycles, updating the new chunk node until it represents the sequencing of phonemes in each syllable and the sequencing of syllables in the word. While chunking phonemes into syllables and syllables into words, the syllable template imposes language-specific constraints on the order of phonemes and syllables. Rehearsal serves the function of refreshing traces held in the phonological store and preventing them from decaying. When weights begin to decay within this model, they can be strengthened via a re-presentation of the nonword/novel word through: a) rehearsal or b) another encounter with the word in the environment. When weights are strengthened new words are learnt permanently.

If PSTM is a language learning device, its mechanisms of learning novel word forms should also be applicable to the L2 context. PSTM should, therefore, play a role in L2 vocabulary, at least at initial stages of L2 learning. Unlike adults, however, monolingual children learning their L1 have only one (developing) linguistic system at their disposal. Adults learning their L2 have access to knowledge of L1 phonological regularities and to a network of L1 semantic information. The Gupta and MacWhinney (1997) model deals with one linguistic system only, and offers no suggestion as to how another set of phonological regularities, as in L2, might interact with the modelled

system. The question arises as to whether the activation of L1 long-term knowledge, in the case of adult Polish learners of English, is sufficient to compensate for individual differences in PSTM functioning. If not, the strength of the link between PSTM and vocabulary knowledge/learning in that sample of learners would be likely to be comparable to that detected in child L1 learning.

The picture emerging from studies involving adults learning L2 in a classroom context is a mixed one. On the one hand, the study by O'Brien et al. (2006) suggests that a similar pattern to that observed in children may be present in adult L2 learners in the classroom context. On the other hand, the study by Speciale et al. (2004: Experiment 2) indicates that there may be no significant link between PSTM and L2 vocabulary learning in L2 adult learners even at the very beginning stage of L2 learning. The difference between the results obtained by O'Brien et al. and those obtained by Speciale et al. may have been brought about by the nature of the L2 vocabulary measures used. O'Brien et al. measured free productive knowledge of L2 vocabulary in spoken output, while Speciale et al. measured the students' ability to arrange some high frequency words into 10 pairs that could be reasonably contrasted. As only certain words can be contrasted in terms of their meaning, it is not unlikely that the Speciale et al. receptive L2 vocabulary task was too restricted in its content to reflect the amount of progress their students made at the duration of the study. In contrast to the Speciale et al. study (Experiment 2), a number of laboratory-based studies involving adults indicate that there is a strong link between PSTM and L2 vocabulary learning at the onset of L2 learning (Craig & Watkins, 1973; Gupta, 2003; Papagno & Vallar, 1992; Speciale et al., 2004, Experiment 1).

Considering the above inconsistencies, the current study explores the link between PSTM and L2 vocabulary knowledge/learning using a PSTM measure that involves articulation (like in the Speciale et al. study, Experiment 2) and a PSTM measure that does not involve articulation (like in the O'Brien et al. study). It also includes L2 vocabulary measures that are of a more general nature (like in the O'Brien et al. study)⁸¹ and L2 vocabulary measures that test the content of the language course

⁸¹ Please note that the O'Brien et al. (2006) measure tapped a different construct from our vocabulary measures in that it measured free productive use of L2 vocabulary in speech, and our measures tapped receptive and controlled productive knowledge of L2 vocabulary in a written format. In the current study and in the O'Brien et al. study, L2 vocabulary was measured at Time 1 and Time 2 to capture aspects of vocabulary knowledge at those times as well as the rate of vocabulary learning between those times.

only (like in the Speciale et al. study, Experiment 2). It explores the issue of the relationship between PSTM and L2 vocabulary knowledge/learning at different L2 proficiency levels, as it has not yet been addressed in respect of adult L2 learners. This study also focuses on the role of PSTM in L2 collocational knowledge/collocation learning. Since only one laboratory-based study has addressed the issue of PSTM and L2 collocation learning in adults (Ellis & Sinclair, 1996; see Chapter 4, Section 4.6 for more detail), we felt it should be further explored in a longitudinal classroom-based design.

The measures used to address Question 1 and Question 4 are as follows. The PSTM capacity was measured by Serial Nonword Recall and Serial Nonword Recognition at the onset of the study (Time 1). L2 vocabulary and collocation learning in class was tapped at the end of the study (Time 2) by the (A2/B1) Progress Vocabulary Test and the (A2/B1) Progress Collocation Test respectively. L2 vocabulary learning outside the classroom during the corresponding period was measured by the difference in performance between Time 1 and Time 2 on the (A2/B1) Vocabulary Levels Test. L2 vocabulary knowledge at Time 1 and 2 was tapped by scores on the (A2/B1) Vocabulary Levels Test (Time 1 & 2). L2 collocation learning outside the classroom was measured by the difference in the performance on the (A2/B1) Gitsaki Collocation Test between Time 1 and Time 2. L2 collocation knowledge at Time 1 and Time 2 was tapped by the (A2/B1) Gitsaki Collocation Test (Time 1 & 2)⁸².

The current study shows that, as is the case of child L1 learning (Gathercole et al., 1992) and child L2 learning (French, 2003), a strong link between PSTM and the knowledge/learning of vocabulary is detectible at the beginning stages of adult L2 learning. In our A2 group, a strong link is detected between PSTM and L2 vocabulary knowledge, as well as between PSTM and L2 vocabulary learning. With regard to L2 vocabulary knowledge, the correlations between PSTM measures (Time 1) and the A2 Vocabulary Levels Test scores (Time 2) are highly significant. With regard to L2 vocabulary learning, significant correlations are observable between PSTM measures

⁸² See Chapters 5 and 6 for a detailed description of relevant research instruments and Chapter 1 for a description of operational definitions of key variables. With regard to two proficiency levels under scrutiny (A2 and B1), please note that, even though the OUP test was used to assign subjects to respective proficiency groups, it is plausible that variability in overall proficiency is not exactly the same in both groups.

(Time 1) and the A2 Vocabulary Levels Test gain scores. The above correlations remain significant when Exposure is partialled out. This indicates that there is a relationship between our subjects' L1 Serial Nonword Recall and Recognition scores, and that there is an increase in knowledge of high-frequency L2 vocabulary over the six-month period. There are also strong correlations between the PSTM measures and the A2 Progress Vocabulary Test, which demonstrates that individual differences in PSTM scores at the commencement of the TCD English language course are linked with corresponding variations in the scores on the end-of-course vocabulary test.

As mentioned earlier, O'Brien et al. (2006) did not identify a significant correlation between their L1-based Nonword Serial Recognition measure and gains in free productive use of L2 vocabulary by adult learners. As a result, they hypothesize that perhaps PSTM "is not related to the development of productive vocabulary [use] but, rather, to the development of vocabulary knowledge" (O'Brien et al., 2006: 395). The current study confirms the latter part of the above statement, namely that the link between Serial Nonword Recognition and gains in high frequency L2 vocabulary is statistically significant in adult A2 learners of English. This study also shows that there is a significant link between Serial Nonword Recall (Time 1) and L2 vocabulary gains in adults.

In the B1 group, the correlations between PSTM and our L2 vocabulary measures are weaker than those obtained in the A2 group. The correlations between Serial Nonword Recall and L2 vocabulary measures mentioned below are significant at the .05 level, but become non-significant when Exposure is partialled out. Serial Nonword Recall is linked with the B1 Vocabulary Levels Test (Time 2) at the .05 level, but is reduced to a non-significant level when Exposure is partialled out. The correlation between Serial Nonword Recall and the B1 Vocabulary Levels gains is significant ($p = .046$), but when Exposure is partialled out the link becomes non-significant. The correlation between Serial Nonword Recall and the B1 Progress Vocabulary Test (Time 2) is significant, but again the correlation becomes non-significant when Exposure is partialled out. The correlation between Serial Nonword Recall and the B1 Progress Vocabulary Test (Time 2) is non-significant. The correlations between Serial Nonword Recognition and all L2 vocabulary measures are all non-significant (all $ps > .05$).

The fact that the correlations between Serial Nonword Recall (Time 1) and L2 vocabulary measures (Time 2 and gain scores) in the B1 group are significant at the .05 level, and that they become non-significant when Exposure is partialled out could be explained in terms of the existence of a high correlation between Serial Nonword Recall (Time 1) and Exposure in the B1 group. Since both Serial Nonword Recall (Time 1) and Exposure correlate with L2 vocabulary measures (Time 2 and gain scores), it is statistically impossible to separate the influence of Serial Nonword Recall (Time 1) and Exposure on L2 vocabulary scores. This issue is dealt with further in Section 8.2.1.

The results obtained in the A2 and B1 adult learners correspond to the finding by Gathercole et al. (1992) that the link between PSTM and L1 vocabulary knowledge decreases as children's L1 proficiency increases. Apart from showing that the involvement of PSTM in L2 vocabulary knowledge/learning decreases as L2 proficiency increases, the current study also shows that there is a link between L1 vocabulary knowledge and L2 vocabulary knowledge at an early stage of L2 learning. In the A2 group, the L1 Vocabulary Test correlates significantly with the A2 Vocabulary Levels Test at both Time 1 and Time 2. In the B1 group, on the other hand, the L1 Vocabulary Test does not correlate with respective L2 vocabulary tests. Even though there is a significant link between L1 vocabulary knowledge and L2 vocabulary knowledge in the A2 group, PSTM measures remain significantly correlated with the A2 Vocabulary Levels Test (Time 1 and Time 2) even when the L1 Vocabulary Test scores are partialled out. This could indicate that at initial stages of L2 learning (A2 level) adult Polish learners of English successfully draw upon some aspects of their L1 vocabulary knowledge, presumably also including L1 phonological regularities. This activation of their L1 long-term knowledge, however, is not sufficient to render the correlation between PSTM and L2 vocabulary measures insignificant. In other words, it is not sufficient to compensate for individual differences in PSTM functioning.

L2 collocation knowledge and learning

The results presented in the previous section demonstrate that PSTM plays a role in long-term learning of novel word forms. Additionally, PSTM may also play a role in the long-term learning of a) the order of novel word forms in a sequence and b) the order of familiar word forms in a sequence. Computational STM models, such as

the Gupta and MacWhinney (1997) model, apart from simulating recall of sequences of nonwords (novel word forms) also simulate recall of sequences of digits (familiar word forms).

We hypothesise that there may be a difference in the degree of PSTM involvement in L2 vocabulary and L2 collocation knowledge/learning at a post-beginner level of L2 proficiency (B1 level in the current study). Since the amount of exposure to a given word is likely to be higher than the amount of exposure to a collocation in which that word features, it may take longer for the links between words comprising two-word collocations to be strengthened than for the links between syllables comprising a word to be strengthened. Additionally, some two-word collocations can be separated by other words on occasion, and this can also contribute to weaker links between their constituent word units. When it comes to learning two-word collocations, the way that words co-occur appears less predictable when compared to sequencing of syllables within a word. Therefore, we speculate that the involvement of PSTM in the learning of L2 collocations at the B1 level of L2 proficiency may be stronger than PSTM involvement in the learning of L2 vocabulary.

The relationship between PSTM and the knowledge/learning of L2 collocations, to the best of our knowledge, has only been explored in a laboratory-based study (Ellis & Sinclair, 1996) that used articulatory suppression to impair PSTM functioning, and did not measure the subjects' variation in PSTM scores. The articulatory suppression condition, which involves the use of a demanding secondary task that competes for the limited capacity of PSTM, was shown to produce a lower number of correctly remembered L2 phrases than other experimental conditions.

The relationship between PSTM and L2 collocation learning/knowledge in the A2 group turns out to be strong. Both Serial Nonword Recall and Serial Nonword Recognition correlate highly with the A2 Gitsaki Collocation Test gain scores and the A2 Progress Collocation Test. The correlation between our two PSTM measures and the A2 Gitsaki Collocation Test (Time 2) is stronger than that between the PSTM measures and the A2 Gitsaki Collocation Test (Time 1). The strength of the link between our Serial Nonword Recognition measure (Time 1) and the A2 Gitsaki Collocation Test (Time 1 and 2) resembles the strength of the correlations between a measure of the use of subordinate clauses and an L1-based Serial Nonword

Recognition task used by O'Brien et al. (2006) in a group of low ability students. This extends the results obtained by Ellis and Sinclair (1996) in a laboratory-based study of the learning of L2 phrases to a real classroom-based setting.

Unlike the results obtained in relation to L2 vocabulary, when Exposure is partialled out, A2 *and* B1 collocation measures are significantly linked to Serial Nonword Recall. A comparison of patterns of correlations obtained in the A2 and B1 groups reveals that the strength of the correlations between our Serial Nonword Recall and L2 collocation measures slightly decreases as L2 proficiency increases. In the A2 group Serial Nonword Recall correlates with the A2 Gitsaki Collocation Test (Time 2) and with the A2 Gitsaki Collocation Test gains at the .01 level. In the B1 group Serial Nonword Recall correlates with the B1 Gitsaki Collocation Test (Time 2), the B1 Gitsaki Collocation Test gains, and the B1 Progress Collocation Test at the .05 level (even when Exposure is partialled out).

When Exposure is partialled out, Serial Nonword Recall in the B1 group no longer correlates with the B1 Vocabulary Levels Test gain scores, but it remains significantly linked with the B1 Gitsaki Collocation Test gain scores. This could be interpreted as support for O'Brien et al.'s (2006: 399) claim that at post-beginning stages of L2 learning, PSTM, instead of being deployed for L2 vocabulary learning, is deployed for learning more complex structures. L2 vocabulary learning at post-elementary levels of L2 proficiency can be aided by existing L2 lexical knowledge in semantic, conceptual and phonological dimensions, and therefore does not need to rely heavily on PSTM (Gathercole, 2006a: 515). L2 collocation learning at the B1 level may require more processing power and a greater learning effort than the learning of single words. The study shows that, although the link between the Serial Nonword Recall task and L2 vocabulary learning diminishes at the B1 level, the link between Serial Nonword Recall and L2 collocation learning remains strong.

When the strength of the correlations between Serial Nonword Recognition and L2 collocation measures is compared in the A2 group and the B1 group, a weakening similar to that observed in relation to L2 vocabulary learning becomes apparent. In the A2 group the correlations between Serial Nonword Recognition and L2 collocation measures are significant at the .05 level. However, in the B1 group the correlations between Serial Nonword Recognition and L2 collocation measures are all non-

significant. This issue of the results obtained via the Serial Nonword Recognition measure is dealt in more detail in Sections 8.1.2 and 8.2.2.

The data obtained in the course of the current project, in our opinion, extend the Ellis (1996) model of L1 collocation learning to L2 adult collocation learning. Ellis takes the line that L1 learning, apart from learning individual words, also involves learning sequences of words (see Chapter 4). This study shows that adult L2 learners engage in the learning of sequences of words; there is a significant increase in the A2 and B1 subjects' collocational knowledge between Time 1 and Time 2. The current study also *partially* supports the Wray (2002) model of L2 collocation learning. Wray claims that adults tend to break up newly encountered phrases into words, and that the information about how these words are combined is not retained. In the low PSTM subgroups in the A2 and B1 groups the difference between Time 1 collocation scores and Time 2 collocation scores is not statistically significant, which indicates that *some* adults (namely adults with lower PSTM functioning) may be less efficient at chunking at a lexical level and may not easily retain information about word combinations. Our data do not support the claim that adults in general fail to retain the information about how words are combined.

In the light of the findings regarding bilingual L2 learners' ability to chunk at a word level and at a collocation level, it would seem advisable to broaden the scope of such models as the Gupta and MacWhinney (1997) STM model in order to better reflect the functioning of the phonological loop. Burgess and Hitch (2005: 539) suggest that a more realistic computational model should contain "a hierarchy of context signals [that] deals with ordering at different levels of representation, from phonemes through words to groups or lists". Such a hierarchy of context signals could enable modelling of STM and long-term learning at the levels of syllables, word forms, and sequences of words, including the learnt familiarity of words and sequences of words. Ideally, this model should also include an interaction between one linguistic system (L1) and another linguistic system (L2) during long-term learning of words and chunks.

8.1.2. Sub-components of PSTM, L2 vocabulary/collocations and adult L2 proficiency

Question 2⁸³ and Question 5⁸⁴ address the relationship between two subcomponents of PSTM, the phonological store and the articulatory rehearsal component, and L2 vocabulary/collocation learning at the A2 and B1 levels of L2 proficiency.

Within the Working Memory model (1974) the Serial Nonword Recognition task is believed to mainly involve the nonarticulatory phonological store, while the Nonword Repetition and Serial Nonword Recall tasks are assumed to involve both the store and the articulatory rehearsal component (Baddeley, 2003a). The existence of the distinction between the phonological store and the rehearsal component is supported by neuropsychological evidence from patients who experience disrupted storage but not rehearsal, or the other way round (Howard & Franklin, 1990; Vallar & Papagno, 2002). In neuroimaging studies the phonological storage component has been located in the temporo-parietal region in the left hemisphere, while the rehearsal component has been located more frontally in Broca's area (Jonides et al., 1998; Paulesu et al., 1993). Normal adult subjects, such as those participating in the current study, would not be expected to have a serious impairment of one or the other component. They would, though, be expected to exhibit some variation both in storage capacity and rehearsal efficiency.

Since recall-based nonword tasks have been used much more often in PSTM research than recognition-based nonword tasks have, it is important to explore whether in adult L2 learners it is the phonological store (tapped by Serial Nonword Recognition) that is more closely linked to the learning of L2 vocabulary/ collocation, or whether it is both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall). In the case of young children the link between L1 vocabulary and a recognition-based nonword measure was shown to be stronger than

⁸³ **Question 2:** Which component of PSTM is more closely linked to L2 vocabulary learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?

⁸⁴ **Question 5:** Which component of PSTM is more closely linked to L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency – the phonological store (tapped by Serial Nonword Recognition) or both the articulatory rehearsal component and the phonological store (tapped by Serial Nonword Recall)?

that between L1 vocabulary and a recall-based PSTM measure (Gathercole et al., 1999b). In an experimental study of adult artificial language learning by Martin (2009: 87 & 90), on the other hand, the strength of the correlation between some aspects of L2 vocabulary learning (and L2 word order) and scores on a recall-based PSTM measure ($p < .05$) was stronger than between these same aspects and a recognition-based PSTM measure ($p > .05$). This difference between young children and adults could be related to the finding that adults make use of rehearsal strategies in order to hold novel items in memory (Pressley & Van Meter, 1992), while young children do not (Gathercole & Hitch, 1993; Pickering, 2006: 245). Adults are able to make use of strategic rehearsal, such as for example cumulative or chunked rehearsal, while engaging in immediate serial recall, and therefore we hypothesise that, out of the two PSTM measures, Serial Nonword Recall should produce stronger correlations with scores on the learning of L2 vocabulary/collocation.

The measures used to address Question 2 and Question 5 are as follows. PSTM capacity was measured by Serial Nonword Recall and Serial Nonword Recognition at the onset of the study (Time 1). L2 vocabulary learning outside the classroom was measured by the difference in performance on the (A2/B1) Vocabulary Levels Test between Time 1 and Time 2. L2 collocation learning outside the classroom was measured by the difference in performance on the (A2/B1) Gitsaki Collocation Test between Time 1 and Time 2.

In order to investigate the contribution of the articulatory rehearsal component to predicting the A2 Vocabulary Levels Test gain scores, we carried out two hierarchical multiple regressions. When Serial Nonword Recognition was entered in Block 1, and Serial Nonword Recall was entered in Block 2, Serial Nonword Recall explained an additional 7.2 per cent of variance in the A2 Vocabulary Levels Test gain scores. When the entry order was reversed, Serial Nonword Recognition, entered in Block 2, explained an additional 0.8 per cent of variance. For the purpose of the study, we assume that the amount of additional variance that Serial Nonword Recall accounted for after Serial Nonword Recognition had been added to the model (i.e. 7.2 per cent) can be attributed to the articulatory rehearsal component. The fact that the amount of additional variance explained by Serial Nonword Recall (Time 1) is considerably higher than the amount of additional variance explained by Serial

Nonword Recognition (Time 1) – 7.2 per cent and 0.8 per cent respectively – can be interpreted as an indication that the articulatory rehearsal component contributes to gains in the knowledge of L2 vocabulary in adult learners. Considering the fact that in the current study the amount of variance in the A2 Vocabulary Levels Test gain scores accounted for by the phonological store is 14.7 per cent, we could argue that in adult L2 learning at the A2 level of proficiency both the phonological store and the rehearsal component are actively involved, which is in keeping with the Baddeley and Hitch WM model. The results obtained in relation to the contribution of the phonological store and the articulatory rehearsal component to the A2 Gitsaki Collocation Test gain scores follow the same pattern. When Serial Nonword Recognition was entered in Block 1 and Serial Nonword Recall was entered in Block 2, Serial Nonword Recall explained an additional 7.3 per cent of variance in the dependent variable. When the entry order was reversed, Serial Nonword Recognition entered in Block 2 explained only an additional 1.8 per cent of variance.

The results obtained in the A2 group lend support to Gupta's (2003) suggestion that the strength of the correlations between recall-/recognition-based PSTM tasks and child L1 learning differs from the strength of the correlations between recall-/recognition-based PSTM tasks and adult L2 learning. As hypothesised, Serial Nonword Recall produces stronger correlations with scores on tests measuring the learning of L2 vocabulary/collocations. As mentioned earlier, this difference between young children and adults may be due to the fact that adults engage in strategic rehearsal to strengthen phonological traces, while children do not.

In the analysis of the relationship between PSTM and B1 vocabulary learning, the fact that Exposure correlated significantly with Serial Nonword Recall and with measures of B1 vocabulary knowledge/learning turned out to be somewhat problematic. As pointed out earlier, a high correlation between two independent variables increases the probability that, when one of them is entered into a multiple regression model subsequent to the other, the variable entered second will be found to be non-significant and rejected from the model despite possibly being a good predictor of the outcome. When Exposure was entered in Block 1 and Serial Nonword Recognition in Block 2, Serial Nonword Recognition accounted for an additional 1.8 per cent of variance in the B1 Vocabulary Levels Test gain scores. Serial Nonword

Recall, entered in Block 3, explained very little additional variance (0.3 per cent) as well. When the entry order was reversed, both Serial Nonword Recall, entered in Block 2, and Serial Nonword Recognition, entered in Block 3, accounted for a very small amount of variance in the dependent variable (1.6 and 0.4 per cent respectively). This indicates that when Exposure scores are taken into account, the link between the phonological loop and B1 vocabulary learning is not significant, and neither is the link between the articulatory rehearsal component and B1 vocabulary learning. It seems reasonable to interpret this finding as an indication that at the B1 level of proficiency neither low capacity of the phonological store nor low efficiency of the rehearsal process significantly impairs the learning of L2 vocabulary.

The last set of hierarchical multiple regressions with a different entry order deals with B1 Gitsaki Collocation Test gain data. When the B1 Gitsaki Collocation Test gain scores were entered into a regression model as a dependent variable, Serial Nonword Recognition was entered in Block 1 and Serial Nonword Recall was entered in Block 2, Serial Nonword Recall explained an additional 12.4 per cent of variance in the dependent variable. When the order of entry was reversed, Serial Nonword Recognition, entered in Block 2, explained less than 0.01 per cent of variance. The additional amount of variance in the dependent variable accounted for by the articulatory rehearsal component (Serial Nonword Recall in Block 2) was significant. This could be interpreted as an indication that the lower efficiency of the articulatory rehearsal component constrains the learning of L2 collocations more than it does the learning of L2 vocabulary.

8.1.3. PSTM and adult L2 vocabulary/collocation: causal direction and proficiency

Questions 3⁸⁵ and Question 6⁸⁶ address the issue of the direction of causality of the relationship between PSTM and L2 vocabulary/ collocational knowledge. PSTM was tapped by Serial Nonword Recall and Serial Nonword Recognition (Time 1 and Time 2). L2 vocabulary and collocational knowledge was indexed by the (A2/B1) Vocabulary Levels Test and the (A2/B1) Gitsaki Collocation Test (Time 1 and Time 2).

There are two alternative and opposing accounts of the direction of causality in this relationship, namely the so-called PSTM account (Baddeley et al., 1998) and the linguistic account (Snowling et al., 1991)⁸⁷. The PSTM account posits that PSTM causally determines both Nonword Recall/Recognition ability and vocabulary knowledge/learning, and that the correlation between Nonword Recall/Recognition ability and vocabulary knowledge/learning reflects the effect of PSTM capacity as the underlying variable. The linguistic account (Snowling et al., 1991), on the other hand, proposes that it is the extent of lexical/phonological knowledge that is the causal determinant of Nonword Recall/Recognition ability, and that when lexical/phonological knowledge improves so does the performance on Serial Nonword Recall/Recognition tasks. It should be pointed out that the PSTM account acknowledges that factors other than PSTM can influence the performance on Nonword Repetition, Serial Nonword Recall and Serial Nonword Recognition, and that one such relevant factor is lexical/phonological knowledge of an individual (Gathercole, 2006a; see also the discussion of the notions of lexicality, wordlikeness and redintegration in Chapter 2). The major difference between the two accounts is, therefore, the role ascribed to PSTM in vocabulary knowledge/learning. The PSTM account presents PSTM as causally linked with the growth of vocabulary knowledge, while the linguistic account posits that the involvement of PSTM may be relatively minor.

⁸⁵ **Question 3:** In the case of adults (as opposed to children who were the focus of previous research), what is the causal direction of the relationship between PSTM and the learning of high-frequency L2 vocabulary (outside the classroom) at the A2 and B1 levels of L2 proficiency? Is it the L1-based nonword scores at Time 1 that contribute to L2 Vocabulary Levels Test scores at Time 2, or is it L2 Vocabulary Levels Test scores at Time 1 that contribute to the L1-based nonword scores at Time 2?

⁸⁶ **Question 6:** What is the causal direction of the relationship between PSTM and L2 collocation learning in adults at the A2 and B1 levels of L2 proficiency? Is it PSTM scores at Time 1 that contribute to Gitsaki's Collocation Test scores at Time 2, or is it Gitsaki's Collocation Test scores at Time 1 that contribute to the PSTM scores at Time 2?

⁸⁷ This terminology is adopted from Gupta and Tisdale (2009).

The method used in this study to investigate the issue of causality between PSTM and L2 vocabulary/collocation knowledge in L2 adult learners is a cross-lagged panel correlation paradigm. A cross-lagged correlation analysis involves measuring Variables A and B at Time 1 and Time 2, and comparing the strength of the correlation between Variable A at Time 1 and Variable B at Time 2 with the strength of the correlation between Variable B at Time 1 and Variable A at Time 2. According to the assumptions of a cross-lagged correlational analysis, the correlation should be stronger in the causal than in the noncausal direction (e.g. Campbell, 1988; Campbell & Stanley, 1963; Crano et al., 1972; Crano & Mellon, 1978).

In the A2 group the correlation between Serial Nonword Recall (Time 1) and the A2 Vocabulary Levels Test scores (Time 2) is significant, and so is the converse correlation between the A2 Vocabulary Levels Test scores (Time 1) and Serial Nonword Recall (Time 2). The former correlation is greater than the latter and, as the Pearson and Filon test indicates, the difference between the two correlation coefficients is significant. The correlation between Serial Nonword Recall (Time 1) and the A2 Gitsaki Collocation Test scores (Time 2) is also significantly stronger than the correlation between the A2 Gitsaki Collocation Test scores (Time 1) and Serial Nonword Recall (Time 2). The correlations in the B1 group show a weakening effect of PSTM on L2 vocabulary/collocation knowledge. The correlation between Serial Nonword Recall (Time 1) and the B1 Vocabulary Levels Test (Time 2) is not significantly different from the correlation between the B1 Vocabulary Levels Test (Time 1) and Serial Nonword Recall (Time 2). The same applies to cross-lagged correlations between Serial Nonword Recall and B1 Gitsaki Collocation Test scores.

The pattern of results obtained at the A2 level provides support for the PSTM account, according to which PSTM is a causal determinant of subsequent vocabulary knowledge. Our results are not consistent with the lexical account, according to which the role of PSTM in L2 vocabulary/collocation learning is minimal. This study extends the finding by Gathercole et al. (1992) in relation to PSTM and L1 knowledge in children aged 4 and 5 to early stages of L2 vocabulary and collocation learning in adult L2 learners. This study also supports a recent computational model put forward by Gupta and Tisdale (2009), which demonstrates in a series of simulations that low PSTM capacity (or PSTM functioning to use the authors' terminology) restricts the

speed of vocabulary learning (Simulation 5). In the light of this, Gupta and Tisdale (2009: 496) argue that PSTM functionality is causally linked with vocabulary learning and plays a crucial role in the process.

The pattern of results obtained at the B1 level also appears to be in keeping with the PSTM account, according to which the link between PSTM and vocabulary learning is likely to weaken as language proficiency increases. Our data suggest that at the B1 level of proficiency the cross-lagged correlations between Serial Nonword Recall and L2 vocabulary/collocation knowledge are not significantly different, which can be interpreted as a weakening of the influence of PSTM. As Gathercole (2006a) points out, the more proficient L2 learners are, the more adept they become at lexically supported learning. Instead of relying on PSTM they can access phonological representations of close neighbours of a new L2 word.

The fact that in the B1 group the correlations between the L2 vocabulary/collocation scores (Time 1) and the Serial Nonword Recall scores (Time 2) are significant could be interpreted as an indication of the influence of lexical/phonological knowledge on subsequent PSTM performance – in line with both the PSTM account and the linguistic account. In the literature there are a number of studies indicating that the level of earlier lexical/phonological knowledge in a given language (L_X) influences later performance on L_X -based PSTM tasks, and that performance on L_X -based PSTM tasks improves as L_X proficiency increases (Gathercole, 1995; Gathercole et al, 1991). The Gupta and Tisdale (2009) computational model also indicates that Nonword Repetition ability increases as a function of increasing vocabulary knowledge and increasing knowledge of phonological regularities (Simulation 4). However, even though L2 vocabulary/collocation knowledge (Time 1) is significantly linked with our L1-based Serial Nonword Recall results (Time 2), it does not enhance performance on the Serial Nonword Recall task at Time 2. In the current study the Serial Nonword Recall task is not L2-based, and therefore the impact of L2 lexical/phonological knowledge (Time 1) on performance on L1-based Serial Nonword Recall (Time 2) should be expected to be minimal. We suspect that a significant correlation between L2 vocabulary/collocation knowledge (Time 1) and Serial Nonword Recall (Time 2) arises out of the fact that the Serial Nonword Recall scores (Time 1) and the Serial Nonword Recall scores (Time 2) are very highly

correlated. Serial Nonword Recall scores at Time 1 and Time 2 correlate significantly with L2 vocabulary/collocation scores at Time 1, and our subjects' scores on the L1-based Serial Nonword Recall task remain stable over time.

To recapitulate the above points, at the A2 level of L2 proficiency PSTM is causally linked with L2 vocabulary knowledge/learning (Figure 8.1 below). The link between our L1-based Serial Nonword Recall and L2 vocabulary knowledge reflects the effect of PSTM capacity as the underlying variable. The same applies to the relationship between PSTM (as tapped by L1-based Serial Nonword Recall) and L2 collocational knowledge. At the B1 level of L2 proficiency the relationship between L1-based Serial Nonword Recall and L2 vocabulary weakens, and becomes non-significant when exposure to L2 outside the classroom is partialled out. The relationship between the Serial Nonword Recall task and L2 collocation learning/knowledge remains significant at the B1 level but noticeably weaker from the link between the two variables at the A2 level. Additionally, the current study shows that L2 vocabulary knowledge does not cause gains in the performance on L1-based Serial Nonword Recall, and that when L1-based nonword of low wordlikeness are used in adults' performance on Serial Nonword Recall remain stable over time.

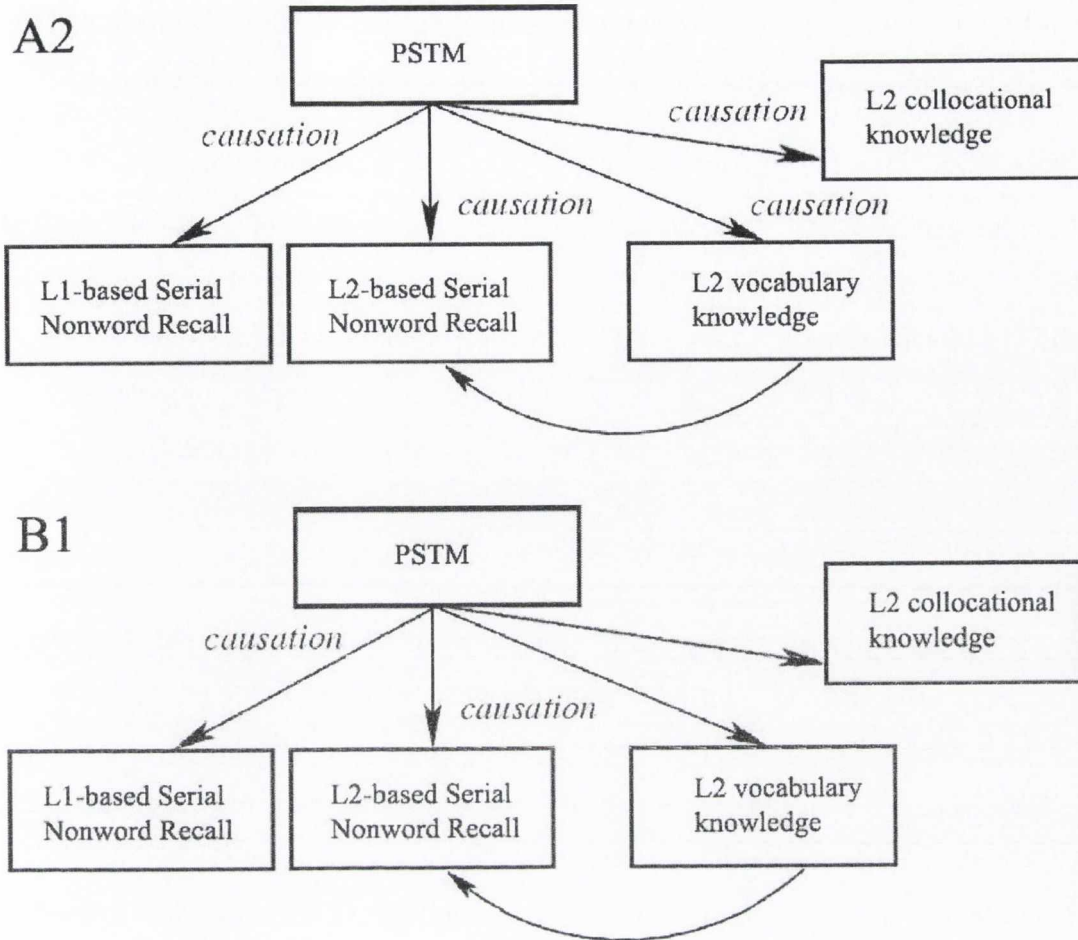


Figure 8.1. Depiction of the relationship between Serial Nonword Recall and L2 vocabulary/collocation knowledge/learning in adults at A2 and B1 levels of L2 proficiency

The patterns of results obtained by using of the Serial Nonword Recognition task differs from the above figure in that no significant differences between pairs of cross-lagged correlations were discovered at the A2 level or the B1 level. There are at least two possible explanations why no significant differences between pairs of cross-lagged correlations were discovered at the A2 level in relation to the Serial Nonword Recognition scores, while significant differences were detected in relation to the Serial Nonword Recall scores. First, since our subjects are *normal* adults who have access to both PSTM components, it is only when a PSTM measure that taps *both* components is used that PSTM emerges as a causal determiner of subsequent L2 vocabulary/collocation knowledge at the A2 level. Second, it is also plausible that the correlations involving the Serial Nonword Recognition measure would have turned out to be stronger, had we had access to a larger sample. Nimmo (2004) suggests that Serial

Nonword Recognition may require a larger sample size than Serial Nonword Recall to yield generalizable results. This, however, has to yet be confirmed empirically.

Additionally, it is also possible that in adult subjects Serial Nonword Recognition does not tap 'pure' phonological store capacity without the involvement of rehearsal. The results obtained on the Serial Nonword Recognition measure raise a question as to what exactly is encoded while an individual is engaging in the recognition task, and how much this encoding is affected by different compensatory strategies (e.g. mnemonic strategies) used by individuals (see Section 8.2.2).

8.2. Limitations of the present study

We turn now to a consideration of some of the limitations of the experimental design and of the research instruments in the present study. The shortcomings of the design discussed in Section 8.2.1 below include the presence of potentially interfering variables and the issue of the drop-out over the time span of the project. Section 8.2.2 focuses on the impact that the involvement of speech output skills in the Serial Nonword Recall task and lack of a speech output requirement in the Serial Nonword Recognition task may have on PSTM scores obtained by adult subjects. This section also re-capitulates the measures that were taken during the current study to prevent lowering or boosting PSTM scores.

8.2.1. Limitations of a classroom-based longitudinal design

A positive aspect of a classroom-based design is that it approximates the real-life situation under investigation. However, this happens at the cost of having to deal with some potentially interfering variables which are beyond the control of the researcher. In natural environments, such as real-life classroom learning situations, an attempt is normally made to identify and partial out potentially interfering variables. It is fairly obvious, however, that although it may be possible to identify some potentially interfering variables in a classroom-based design, it may not always be possible to adequately control them. In the current study, for example, the level of exposure to L2 outside the classroom in the B1 group correlated so highly with the Serial Nonword Recall scores that it became impossible to disentangle the impact that both variables had on L2 vocabulary learning. In practice this meant that B1 students with lower

Serial Nonword Recall scores received a significantly lower amount of exposure to L2 outside the classroom than B1 students with higher Serial Nonword Recall scores. Since the study was not laboratory-based, we could not have prevented this problem from arising. We had no control over the amount of L2 that our subjects were exposed to outside the classroom, and we could not have anticipated that scores on the Exposure measure would correlate highly with scores on the Serial Nonword Recall measure. We are also aware that in some vocabulary/collocation tasks such as e.g. multiple-choice questions or fill-in-the-blank sentences high-risk takers can score more points by simply venturing to guess when not sure which answer is correct. In the current study we did not control for risk-taking.

Longitudinal studies allow the observation of change in selected dependent variables by performing repeated testing at specified points in time, but they carry an increased risk of subjects dropping out of the study. At the onset of the current study we gathered PSTM, vocabulary and collocation data from over 100 subjects. By the time the TCD course terminated, approximately 30 per cent of our subjects had dropped out. Some of them moved out of Dublin. Others had had their working hours changed and could no longer attend the course. Additionally, a number of subjects had to be excluded from our analyses as a result of reported hearing problems or dyslexia. This left us with 60 subjects altogether (30 per each proficiency group). A sample size of 30 allows the identification of large effects only (Field, 2005). Within the PSTM tradition there are a number of studies that are based on a much smaller sample size than ours (e.g. Gathercole et al., 1999b: study 1; Papagno & Vallar, 1995). Ideally, however, a larger sample size should be used to enable the identification of smaller effects as well.

8.2.2. Limitations of the Serial Nonword Recall and the Serial Nonword Recognition tasks

This section touches upon the issue of the involvement of speech output skills in Serial Nonword Recall and lack thereof in Serial Nonword Recognition. The Serial Nonword Recall task and the Serial Nonword Recognition task resemble one another in that they both involve auditorily presented lists of nonwords, but they differ in relation to the type of responses that the participants are requested to make while completing the two tasks. The Serial Nonword Recall task requires the repeating each list of

nonwords out loud immediately after its presentation, maintaining the order in which nonwords were originally presented. Serial Nonword Recognition, on the other hand, requires indicating whether the order of nonwords within two presented lists is the same or different. Some benefits and downsides of the involvement of speech output skills and lack of thereof in PSTM tasks are related below to the context of the present study.

The involvement of speech output skills in Serial Nonword Recall may be viewed as one of its benefits, but it can also be one of its disadvantages. The Serial Nonword Recall task provides an indication as to whether subjects are able to *remember* and *repeat back* lists of nonwords, with all constituent nonwords maintaining the same order in which they were originally presented. One of the benefits of this articulatory requirement is that it provides researchers with more detailed data, and allows them to analyse the results using different scoring methods (a correct-in-position criterion versus an item recall criterion)⁸⁸. One potential downside to using any recall-based PSTM measure, including Serial Nonword Recall, is that, since speech output skills are involved, obtained scores may be lower than the actual PSTM capacity as a result of (even minor) articulation problems (Gathercole et al., 2001). This is of particular concern in developmental studies of young children, who may differ in their articulatory skills, but obviously this can also have a bearing on the Serial Nonword Recall scores obtained by adult subjects provided they have some articulation problems. Prior to the commencement of our research project our subjects reported no diagnosed speech problems. No articulation problems were detected during one-to-one sessions that our subjects were requested to attend prior to the administration of the Serial Nonword Recall task (Time 1) either. In the light of the above, we are confident that our Serial Nonword Recall task accurately measured PSTM capacity of our adult subjects.

The fact that Serial Nonword Recognition does not require the involvement of speech output skills has a positive and negative side to it. Serial Nonword Recognition provides a measure that separates the phonological store and the participants' speech production abilities (Baddeley, 2003a; Gathercole et al., 2001). Since the Serial

⁸⁸ This is particularly useful in experimental manipulations (see e.g. Gathercole et al., 2001 or Nimmo, 2004). In classroom-based studies like ours the correct-in-position criterion is normally used (see e.g. O'Brien et al., 2006), according to which a participant's response is scored as correct when both the nonwords and their order are recalled correctly.

Nonword Recognition task does not involve a verbal recall of presented lists of nonwords, it is not susceptible to any articulation problems that participants may have. On the other hand, though, it may not always be clear whether the subjects encode the entire nonwords or only parts of presented nonwords. Encoding only *parts* of presented nonwords, e.g. the initial consonant or syllable, would inevitably result in substantial lowering of the burden placed on STM, and would therefore boost the subjects' scores.

The piloting of the Serial Nonword Recognition task alerted us to the fact that the above-mentioned encoding problem may arise when the measure is employed to tap PSTM capacity in adults (Chapter 5, Section 5.3.1). At the piloting stage two subjects, by their own admission, tried to lower the memory burden by holding only each initial syllable in their memory instead of the whole word. To counteract the problem of subjects attempting to encode initial syllables or consonants only, in the main study they were informed that after completing the Serial Nonword Recognition task they would also be asked a) to complete a cued-recall task to assess how many syllables from each nonword they could retrieve from memory, and b) to describe the strategies they used to remember the nonword lists. These precautionary steps appear to have worked in that the Serial Nonword Recognition task was highly correlated with the Serial Nonword Recall task in the main study, and that the Serial Nonword Recognition task replicated some trends obtained in earlier studies employing this measure (e.g. O'Brien et al., 2006). However, it should also be pointed out that, in spite of the precautions described above, the Serial Nonword Recognition task produced significantly lower reliability coefficients than the Serial Nonword Recall task did (see Chapter 7).

8.3. Suggestions for future research

This section looks at a number of areas related to the present study that seem to merit further investigation.

8.3.1. The impact of phonological proximity between L1 and L2 on the relationship between PSTM and L2 learning

The relationship between PSTM and vocabulary learning is now well-established in the early stages of language proficiency – in relation to child L1 learning

(Gathercole et al., 1992), child L2 learning (e.g. French, 2003), and also adult L2 learning (the current study). The contribution of L1 vocabulary knowledge to L2 vocabulary learning observed in the current study highlights the need for an investigation of the impact that the phonological proximity between L1 and L2 can have on the relationship between PSTM and L2 learning. Groups of adults from various linguistic backgrounds (e.g. Czech, English, Chinese, etc.) who learn one L2 (e.g. Polish) could be compared to establish whether, in cases where there is significant overlap of L1 and L2 systems (e.g. Czech and Polish), it may be possible for L1 phonological knowledge to reduce the burden placed on PSTM at the beginning stages of L2 learning.

8.3.2. Types of nonwords used in PSTM tasks

An additional avenue of research would be to investigate is the impact of the type of nonwords employed in PSTM tasks (i.e. L1-based, L2-based, neither L1-based nor L2-based) on the predictions made by the relevant tasks in relation to L2 vocabulary/collocation learning in adult learners. In our longitudinal study we opted for PSTM tasks that were composed of L1-based rather than L2-based nonwords in an attempt to provide L2-independent PSTM scores. Our subjects' scores on the L1-based Serial Nonword Recall and the L1-based Serial Nonword Recognition tasks remained stable over time, which replicates the findings obtained in a recent longitudinal study of children (French & O'Brien, 2008). In the French and O'Brien study performance on a non-L2-based PSTM measure (Arabic Nonword Repetition) did not increase over time in spite of significant gains in L2 knowledge. However, in line with the redintegration hypothesis (Hulme et al., 1991) according to which partially degraded phonological traces are reconstructed on the basis of long-term knowledge (see Chapter 2), performance on an L2-based PSTM measure (English Nonword Repetition) increased significantly alongside the children's developing L2 knowledge. As a result, French and O'Brien conclude that their neither L1- nor L2-based (Arabic) Nonword Repetition task "provided a very much language-free measure of phonological memory capacity" (2008: 480), while their L2-based Nonword Repetition was amendable to influences from existing L2 knowledge. A recent longitudinal study involving adult learners (O'Brien et al., 2006) also shows that in adult L2 learners L1-based Serial Nonword Recognition scores do not increase significantly over time.

In the light of the above, it would be interesting to establish whether the performance on L1-based nonword tasks and the performance on neither L1- nor L2-based nonword tasks among adult L2 learners remain stable over time, and whether the two PSTM tasks yield comparable results in longitudinal designs. Since the phonological storage capacity is thought to be largely language-independent (Baddeley, 2003a; Baddeley, 2007; French & O'Brien, 2008; Gathercole, 2006a), it would be particularly useful to establish which types of nonwords tap this language-independent capacity in a most accurate manner in the adult population.

8.3.3. Serial Nonword Recognition and the construct it measures in adults

The relatively low reliability of the Serial Nonword Recognition task in adults, as compared to the high reliability of the Serial Nonword Recall task, brings up the question of whether Serial Nonword Recognition taps the same processes in adults as it is believed to tap in children. Serial Nonword Recognition has been explored mainly in studies on child L1 learning (e.g. Gathercole et al., 1999b), and therefore more research should be carried out on processes involved in Serial Nonword Recognition in adults. To the best of our knowledge, the Serial Nonword Recognition task has only been used in one other longitudinal study of adult L2 learning apart from ours (O'Brien et al., 2006), but the study in question did not address the above issue. Adults have been shown to employ test-taking and rehearsal strategies that young children do not make use of, which may have implications for what the Serial Nonword Recognition task measures in adults as compared to children. Therefore, we feel it is essential that the issue of the construct validity of the Serial Nonword Recognition task when used in an adult population would be explored in future studies. In the present study Serial Nonword Recall and Serial Nonword Recognition were very highly correlated in both the A2 and B1 groups, which resembles the results obtained in children (Gathercole et al., 2001), but is in contrast with the result obtained in a cross-sectional study of adults by Martin (2009). We believe that the reason why the correlation in the current study was stronger than that reported by Marin is that our subjects received additional instructions for the Serial Nonword Recognition task to discourage them from guessing or engaging in incomplete encoding.

This brings us to a related issue that should be addressed in future research, namely the extent to which different types of instructions affect the performance on the Serial Nonword Recognition task. In the Serial Nonword Recognition task, the participants are normally asked to choose between two options – “same” and “different”. Nimmo (2004: 204) suggests that a slightly different approach be adopted, namely having the participants indicate which nonwords they believe were presented in a different order whenever they choose the “different” option.

8.3.4. PSTM and the second language classroom

Another line of research should focus on applying what has been learnt about the relationship between PSTM and L2 vocabulary/collocation learning in adult learners to classroom techniques and materials in order to enable learners to make the best use of their PSTM capacity. Learners with low PSTM capacity require a higher number of learning trials to remember a new word, and in a small-scale classroom-based study they have been shown to look up words three times as often as students with higher PSTM capacity (Chun & Payne, 2004). It would, therefore, be beneficial to investigate on a larger scale whether providing learners with quick electronic access to definitions and collocations via e.g. multimedia software could speed up their learning process and help them catch up with learners with more efficient PSTM. Another aspect worth investigating is the extent to which visual aids, written props and some input enhancement techniques may compensate for poor PSTM capacity, especially at the beginning stages of L2 learning.

8.3.5. PSTM and free productive use of L2 vocabulary/collocations in adults

Future research on PSTM and L2 vocabulary/collocation learning in adults should also explore free productive use of L2 vocabulary and collocations (see Chapter 3, Section 3.4.3). Students’ scores on language tests such as our L2 vocabulary/collocation tests illustrate only what learners can do in this particular test situation and do not give a full picture of the students’ lexical and collocational knowledge. Student responses elicited via e.g. a fill-in-the-blanks procedure (employed in the (A2/B1) Vocabulary Levels Test, the (A2/B1) Gitsaki Collocation Test, and the (A2/B1) Progress Test) may or may not reflect the kind of language learners produce

spontaneously (Burmeister & Ufert, 1980). They question whether or not the current research findings can be extended to L2 vocabulary/collocation use in samples of students' writing or speech has yet to be investigated.

8.3.6. Construction of new L2 collocation tests

More work should also be done on creating and validating an English collocation test that would probe receptive and productive knowledge of collocations of different types and of different frequencies of occurrence. In spite of some attempts in recent years to create a collocation test for learners of English as L2 (see Chapter 4, Section 4.5.1), currently there is no single English collocation test that would combine the aspects mentioned above. The (A2/B1) Gitsaki Collocation Test we used during our taps into productive knowledge of a wide array of collocations. However, grammatical types of collocations appear to be over-represented, especially those that involve prepositions. The test originated as a progress test in Gitsaki's study (1999) and as such it might have high content validity. When it is used as a general productive collocation test, broader issues arise that involve the representativeness of the target words chosen.

8.4. Conclusion

This study was designed to shed light on the relationship between PSTM (tapped by Serial Nonword Recall and Serial Nonword Recognition) and adult L2 learning. More specifically, the study focuses on the extent to which PSTM functioning affects knowledge and learning of L2 vocabulary and (two-word) collocations in adults. Unlike earlier longitudinal studies dealing with adult L2 learning (O'Brien et al., 2006; Speciale et al., 2004: Experiment 2), this project examines L2 learning at two levels of proficiency (A2 and B1). It also looks at the role of PSTM in the learning of L2 collocations, which has never previously been explored in a longitudinal classroom-based context.

The data collected in the course of the study provide strong evidence of a robust link between PSTM and knowledge/learning of L2 vocabulary/collocations at the beginning stages of adult L2 learning (A2 level). Our data also reveal a weakening of the relationship between PSTM and L2 vocabulary learning/knowledge at a slightly more advanced stage of adult L2 learning (B1 level). In the case of serial recall data,

the link between PSTM and L2 collocations remains significant at the B1 level despite the fact that the link between PSTM and L2 vocabulary learning/knowledge weakens as L2 proficiency increases (and becomes non-significant). This leads us to believe that at the beginning stages of adult L2 learning PSTM functioning has a great impact on the rate of L2 vocabulary learning, but at higher levels of L2 proficiency PSTM is employed for more complex structures such as collocations.

The data reported here go beyond establishing a link between PSTM and aspects of adult L2 learning. They reveal that at the A2 level, PSTM is a *causal* determinant of subsequent knowledge/learning of L2 vocabulary and collocations. At the B1 level, PSTM is no longer causally linked with L2 vocabulary and collocations, which indicates that the direct involvement of PSTM in L2 learning either decreases or disappears as familiarity with the L2 increases. Accumulated knowledge of phonological regularities in L2 becomes an increasingly important factor in L2 vocabulary as L2 proficiency increases. Knowledge of L1 phonological regularities is also shown to be a factor contributing to L2 vocabulary learning/knowledge at the onset of L2 learning.

Given the scope of the evidence reported in the current study, further research in the area that takes account of other components of the Baddeley and Hitch WM model should prove to be fruitful. Finally, it would be advisable to broaden the scope of such models as the Gupta and MacWhinney (1997) STM model in order to better reflect the functioning of the phonological loop when two linguistic systems (L1 and L2) interact.

APPENDIX A: Consent form

You are kindly invited to participate in the Polish Diaspora research project, which is based in Trinity College Dublin. The project is under the general direction of Prof. D.M. Singleton, and my own research is supervised by Dr John Harris. If you would like to participate in the project, please read the following consent form and then proceed with the questionnaire. Please note that your participation is voluntary. You can withdraw at any point in time without any consequences of any kind.

This part of the Polish Diaspora project involves participation in a free English language course organized at Trinity College Dublin. The classes will be held on Saturday mornings and are scheduled to last 90 minutes each. The duration of the course is 6 months. You will not be required to purchase coursebooks or make photocopies of class handouts, as all the materials will be provided by the teacher. Should you be interested in additional study materials, a list of supplementary materials will be available upon request. Detailed course requirements will be specified by the teacher during the introductory meeting.

If you decide to enrol on the course, you will be expected to take a language proficiency test prior to the commencement of the course and a progress test at the end of the course. You will also be asked to complete the following tasks: two tasks involving the memorization of lists of made-up words, each taking 20 minutes to complete, two vocabulary/collocational tests, each taking 90 minutes to complete, two writing tasks, and regular homework assignments.

Any information or data obtained from you during our research will be regarded as strictly confidential. Each questionnaire and language test will have a unique ID number assigned to it to ensure that all your information is kept anonymous. Under no circumstances will the data elicited via our research instruments be available to anyone other than the research team of the project. Although the data in question will be treated

as research findings for presentation and publication, rest assured that your name and other personal details shall not be disclosed in any publications, discussions or correspondence related to this research. Note that you have the right to inspect the results of the tests you took in connection with this research project.

I, (please print) _____, have read the above information and have agreed to participate in this research project. I understand that my participation is voluntary and that I may withdraw from the project any time. I also understand that information from the questionnaire and language tasks will be treated confidentially and that my name or any other personal details will not be revealed.

Signature: _____

Date: _____

APPENDIX B: Background questionnaire

Instructions:

Please answer the following questions by filling in the blanks, ticking the box or circling the number. Please read the questionnaire and sign the consent form if you agree to participate in the project. You may skip any questions you do not feel comfortable answering or that are irrelevant to your experience.

Thank you for your participation.

Part 1: General Information

1. Full name: _____
2. E-mail address: _____
3. Phone number: _____
4. Sex:
 - ☐ Female
 - ☐ Male
5. Date of birth: _____
6. Three most recent jobs in Ireland *(Please list the most recent first)*:
 - 1) _____
 - 2) _____
 - 3) _____
7. The last three jobs in Poland *(Please list the most recent first)*:
 - 1) _____
 - 2) _____
 - 3) _____
8. Length of residence in Ireland: _____ years _____ months

9. Level of education attained:

- ☐ primary
- ☐ secondary
- ☐ vocational
- ☐ tertiary

10. Educational qualifications:

11. Plans for educational development:

- ☐ I am not intending to do any degrees / any courses in the nearest future
- ☐ I am planning to do a further degree, i.e. _____
- ☐ I am currently enrolled on a degree course, i.e. _____ (degree) in
_____ (institution)

12. Your first language (the language that you spoke first as a child)

- ☐ Polish
- ☐ Polish and another language: _____
- ☐ Other: _____

13. Other foreign languages besides of English *(Please indicate all the foreign languages you have ever learned regardless of your present proficiency in them):*

- ☐ German
- ☐ French
- ☐ Russian
- ☐ Other: _____

14. Intended length of stay in Ireland:

- ☐ for good – I intend to stay in Ireland permanently
- ☐ for a couple of years
- ☐ for up to a year
- ☐ for less than 6 months
- ☐ I am not quite sure yet

Part 2: General English Language Learning Experience and Patterns of Use

15. At what age did you start learning English? _____ years old.

16. In which context did you first learn English?

- ☐ at home
- ☐ in a nursery school
- ☐ at primary school
- ☐ at secondary school
- ☐ at university/college
- ☐ at work
- ☐ other: _____

17. How would you assess you English language proficiency?

- ☐ A1 (beginners)
- ☐ A2 (elementary)
- ☐ B1 (lower-intermediate)
- ☐ B2 (upper-intermediate)
- ☐ C1 (lower-advanced)
- ☐ C2 (advanced)

18. How would you describe your ability in the following skill areas in English?

Skill areas	1	2	3	4	5	NA
	Excellent / near-native	Good	Average	Basic	Poor	Not applicable
Speaking	1	2	3	4	5	NA
Listening	1	2	3	4	5	NA
Reading	1	2	3	4	5	NA
Writing	1	2	3	4	5	NA
Grammar	1	2	3	4	5	NA
Vocabulary	1	2	3	4	5	NA
Pronunciation	1	2	3	4	5	NA
Translation	1	2	3	4	5	NA

19. Which areas of learning English do you feel you most need to improve?

Skill areas	1	2	3	4	5	NA
	Not very important to me				Very important to me	Not applicable
Speaking	1	2	3	4	5	NA
Listening	1	2	3	4	5	NA
Reading	1	2	3	4	5	NA
Writing	1	2	3	4	5	NA
Grammar	1	2	3	4	5	NA
Vocabulary	1	2	3	4	5	NA
Pronunciation	1	2	3	4	5	NA
Translation	1	2	3	4	5	NA

20. How easy do you find it to learn English?

- ☐ very easy
- ☐ quite easy
- ☐ neither easy nor difficult
- ☐ quite difficult
- ☐ very difficult

21. How easy do you find learning the following skills in English?

Skill areas	1	2	3	4	5	NA
	Very easy				Very difficult	Not applicable
Speaking	1	2	3	4	5	NA
Listening	1	2	3	4	5	NA
Reading	1	2	3	4	5	NA
Writing	1	2	3	4	5	NA
Grammar	1	2	3	4	5	NA
Vocabulary	1	2	3	4	5	NA
Pronunciation	1	2	3	4	5	NA
Translation	1	2	3	4	5	NA

22. Which of the following statements applies to you most? *(Tick one answer only)*

I want to study English:

- ☐ to cater only for my everyday needs, e.g. at the post office
- ☐ to be able to express myself in most situations
- ☐ to be able to fully express myself in English

23. What are your other goals in studying English? *(Feel free to tick as many boxes as applicable).*

- ☐ to get hired for a better job / to get promoted in your company
- ☐ to study at an educational institution in an English-speaking country
- ☐ to make friends with English-speaking people
- ☐ to integrate with Irish society
- ☐ to get a better insight into Irish culture
- ☐ other: _____

24. When you speak English, how often is it the case that you speak with the following types of speakers of English?

Speakers of English	Very often	Often	Sometimes	Hardly ever	Never
English native speakers	1	2	3	4	5
Non-native speakers of English	1	2	3	4	5

25. What is the balance of Polish and English in your everyday life in Ireland?

- ☐ much more Polish
- ☐ more Polish
- ☐ more English
- ☐ much more English

26. What is the balance of Polish and English in your life in the following contexts?

Use of Polish and English in Ireland	Much more Polish	More Polish	More English	Much more English	Not applicable
with family	1	2	3	4	NA
with friends	1	2	3	4	NA
at work	1	2	3	4	NA
in public places (banks, shops, etc.)	1	2	3	4	NA
other:	1	2	3	4	NA

27. Please add below any further comments you may have about any of the issues raised in the questionnaire or about the questionnaire itself:

*You have now reached the end of the questionnaire.
Thank you very much for your participation.*

APPENDIX C: Serial Nonword Recall task

2-item lists

1. jestaby – fezbino
2. wskane – biznale
3. foksela – wycmosa
4. bywlera – raplido
5. kosmuca – lagroki
6. mudrygo – jukloty
7. dokrapu – roblewy
8. fezbino – seprody
9. letrumi – gaflosy
10. neglika – ceptyko
11. jukloty – foksela
12. roblewy – jestaby
13. biznale – gibruta
14. seprody – dokrapu
15. gaflosy – neglika

3-item lists

16. wycmosa – bywlera – ceptyko
17. seprody – kosmuca – letrumi
18. jestaby – gibruta – dokrapu
19. biznale – neglika – lagroki
20. jukloty – foksela – raplido
21. ceptyko – jestaby – seprody
22. bywlera – gaflosy – roblewy
23. wskane – fezbino – mudrygo
24. gibruta – lagroki – jukloty
25. neglika – wskane – biznale
26. letrumi – ceptyko – gaflosy
27. mudrygo – wycmosa – bywlera
28. kosmuca – raplido – foksela
29. roblewy – dokrapu – wskane
30. fezbino – letrumi – gibruta

4-item lists

31. roblewy – dokrapu – foksela – jestaby
32. letrumi – wycmosa – gibruta – bywlera
33. seprody – neglika – gaflosy – fezbino
34. ceptyko – wskane – lagroki – kosmuca
35. mudrygo – jukloty – dokrapu – raplido

36. wycmosa – lagroki – fezbino – neglika
37. jestaby – gaflosy – raplido – biznale
38. foksela – biznale – jestaby – gibruta
39. bywlera – ceptyko – letrumi – mudrygo
40. woskane – kosmuca – seprody – roblewy
41. dokrapu – gibruta – neglika – jukloty
42. biznale – letrumi – kosmuca – fezbino
43. mudrygo – roblewy – bywlera – ceptyko
44. gaflosy – raplido – jukloty – wycmosa
45. lagroki – foksela – woskane – seprody

APPENDIX D: Serial Nonword Recognition task

4-item lists

1. pegwika – ruspale – jubleny – kacmoba (S)⁸⁹
pegwika – ruspale – jubleny – kacmoba
2. bostagi – sopryla – gubrawi – fimrosy (D)⁹⁰
bostagi – gubrawi – sopryla – fimrosy
3. gadlipu – zudrami – jedmuda – rysnudo (D)
gadlipu – jedmuda – zudrami – rysnudo
4. nazdoty – bafroce – toskaby – meksona (S)
nazdoty – bafroce – toskaby – meksona
5. lizmato – ceplira – wekluga – dagryko (D)
lizmato – wekluga – ceplira – dagryko
6. sopryla – nazdoty – meksona – ruspale (S)
sopryla – nazdoty – meksona – ruspale
7. gubrawi – bostagi – rysnudo – jubleny (S)
gubrawi – bostagi – rysnudo – jubleny
8. jedmuda – gadlipu – lizmato – ceplira (D)
jedmuda – lizmato – gadlipu – ceplira
9. zudrami – kacmoba – fimrosy – bafroce (D)
zudrami – fimrosy – kacmoba – bafroce
10. wekluga – dagryko – pegwika – toskaby (S)
wekluga – dagryko – pegwika – toskaby

5-item lists

11. wekluga – nazdoty – sopryla – bafroce – zudrami (D)
wekluga – sopryla – nazdoty – bafroce – zudrami
12. kacmoba – fimrosy – pegwika – toskaby – ruspale (S)
kacmoba – fimrosy – pegwika – toskaby – ruspale
13. bostagi – dagryko – jubleny – gubrawi – meksona (S)
bostagi – dagryko – jubleny – gubrawi – meksona
14. rysnudo – jedmuda – gadlipu – lizmato – ceplira (D)
rysnudo – jedmuda – lizmato – gadlipu – ceplira
15. fimrosy – wekluga – pegwika – kacmoba – toskaby (S)
fimrosy – wekluga – pegwika – kacmoba – toskaby

⁸⁹ “S” indicates that the order in which the nonwords were presented in both lists was the same.

⁹⁰ “D” indicates that the two lists were different in terms of the order in which the nonwords were presented.

16. zudrami – bafroce – dagryko – ceplira – sopryla (D)
zudrami – bafroce – ceplira – dagryko – sopryla
17. jubleny – toskaby – nazdoty – ruspale – pegwika (D)
jubleny – toskaby – ruspale – nazdoty – pegwika
18. meksona – lizmato – gubrawi – jedmuda – fimrosy (S)
meksona – lizmato – gubrawi – jedmuda – fimrosy
19. lizmato – bostagi – jedmuda – zudrami – gubrawi (S)
lizmato – bostagi – jedmuda – zudrami – gubrawi
20. gadlipu – rysnudo – meksona – sopryla – dagryko (D)
gadlipu – meksona – rysnudo – sopryla – dagryko

6-item lists

21. dagryko – rysnudo – zudrami – kacmoba – bostagi – wekluga (D)
dagryko – zudrami – rysnudo – kacmoba – bostagi – wekluga
22. bafroce – ceplira – fimrosy – lizmato – ruspale – jubleny (S)
bafroce – ceplira – fimrosy – lizmato – ruspale – jubleny
23. nazdoty – sopryla – pegwika – gubrawi – toskaby – jedmuda (D)
nazdoty – sopryla – gubrawi – pegwika – toskaby – jedmuda
24. gadlipu – meksona – wekluga – toskaby – kacmoba – bostagi (D)
gadlipu – wekluga – meksona – toskaby – kacmoba – bostagi
25. ruspale – gubrawi – jedmuda – zudrami – gadlipu – meksona (S)
ruspale – gubrawi – jedmuda – zudrami – gadlipu – meksona
26. pegwika – jubleny – bafroce – nazdoty – lizmato – sopryla (S)
pegwika – jubleny – bafroce – nazdoty – lizmato – sopryla
27. rysnudo – dagryko – ceplira – wekluga – fimrosy – kacmoba (S)
rysnuo – dagryko – ceplira – wekluga – fimrosy – kacmoba
28. bafroce – gadlipu – rysnudo – pegwika – nazdoty – ceplira (D)
bafroce – gadlipu – pegwika – rysnudo – nazdoty – ceplira
29. meksona – sopryla – bostagi – jubleny – gubrawi – lizmato (S)
meksona – sopryla – bostagi – jubleny – gubrawi – lizmato
30. fimrosy – jedmuda – dagryko – ruspale – zudrami – toskaby (D)
fimrosy – jedmuda – ruspale – dagryko – zudrami – toskaby

APPENDIX E: A2 and B1 Gitsaki Collocation Tests

Instructions:⁹¹

In each of the sentences there is ONE word missing. Please fill in each blank with a word that best completes each sentence.

Test for beginners (A2) (Gitsaki, 1999)

1. She is good _____ typing. (at)
2. Oh, dear! 4 o'clock already. Time _____ when you are having fun! (flies)
3. I _____ stamps and old coins. It's my favourite hobby. (collect)
4. She is a good student. She always _____ her homework. (does)
5. The park is crowded _____ school children for the Sports Day. (with)
6. You need a holiday. Why don't you _____ a trip? (take, make)
7. You can't _____ a diagnosis. You are not a doctor. (make, reach, establish)
8. This is a photo of my family. The girl _____ my father's right is my sister. (on)
9. I've had two _____ of bread and I feel full. I can't eat any more. (points)⁹²
10. Read Helen's note and find _____ what happened to her. (out)
11. He's a good brother. He always looks _____ his baby sister. (after)
12. Look _____ my photos. Which one do you like? (at)
13. I need some stamps. Is there a _____ office around here? (post)
14. Good students _____ notes during classes. (take, make)
15. Take the bus from this bus-stop and get _____ at Tower Street. (off)
16. Help Alexandra to fill _____ the questionnaire. (in, out)
17. Melina is studying French _____ a French University. (at)
18. It is Lisa's birthday. Let's think and _____ some suggestions for her birthday party. (make)
19. I play golf _____ Sundays. (on)

⁹¹ Instructions were provided in Polish and English.

⁹² The phrase "points of the compass" in Sentence 9 ("North, South, East and West are the four _____ of the compass") turned out to be far too difficult for A2 learners at the piloting stage. We decided to replace it by a more level-appropriate phrase – "a slice of bread" (see Chapter 5, Section 5.2.1).

⁹³ Sentence 18 ("It is Lisa's birthday. _____ some suggestions for her birthday party") was re-phrased, as it sounded very much like a classroom instruction for a role-play (see Chapter 5, Section 5.2.1).

20. Here's my _____ number. Please call me tonight. (phone, telephone, mobile)
21. You are a tourist in Greece. Complain _____ the trains. They are always late. (about)
22. I can't speak. I've got a _____ throat. (sore, bad)
23. I called the police a few minutes ago. They are _____ their way. (on)
24. Listen carefully. Write _____ what the people are saying. (down)
25. Ask _____ help from your father. I'm too busy. (for)⁹⁴
26. She is busy _____ the moment. Please call later. (at)
27. John is in Greece now. He is _____ a small Greek island, Samos. (on)
28. Every morning I get _____ early. (up)
29. Poland is north _____ the Czech Republic. (of)⁹⁵
30. I am a _____ of a climbing club. We are going on a climbing trip next week. (member)⁹⁶
31. Hold _____ a minute. How do you spell your name again? (on)
32. Keep young children _____ from the fire. (away)
33. Denise is asking her new manager _____ her salary. (about)⁹⁷
34. I am looking _____ Grafton Street. Do you know where it is? (for)⁹⁸
35. Take this _____ of advice from me. Never lie to your friends. (piece, bit, word)
36. She is opening her bag. She is taking something _____. (out)
37. I always call my friend _____ the evening. (in)
38. He is famous _____ his songs. (for)
39. This is a long story. Work _____ it and make it shorter. (on)⁹⁹
40. I sometimes help my mother _____ the dinner. (with)
41. My father is a farmer. He works _____ every day. (hard)
42. Don't go. Your favourite program is _____ now. (on)
43. The school _____ comes at 7.06 every morning and takes us to school. (bus)¹⁰⁰
44. There are a lot of thieves around. Watch _____ at night. (out)
45. The meeting went _____ for two hours. (on)
46. I can see her. She is playing with the dog _____ the garden. (in)
47. I like classical music. I relax _____ it. (with)

⁹⁴ Sentence 25 ("You are painting your room. Ask _____ help from your father") was re-phrased to make the context more natural and less classroom-like.

⁹⁵ More familiar references to Poland and the Czech Republic were added.

⁹⁶ Sentence 30 ("You are a _____ of a climbing club. Plan your next climbing trip") was re-phrased for the same reason as in the case of Sentence 25.

⁹⁷ Sentence 33 ("You are talking to your new manager. Ask _____ your salary") was re-phrased for the same reason as in the case of Sentence 25.

⁹⁸ A more familiar reference to Grafton Street was added.

⁹⁹ "This is the story" was replaced with "This is a very long story" to make it more coherent with the sentence that follows.

¹⁰⁰ Sentence 43 ("Every morning the school bus comes _____ 7.06") was replaced by "The school _____ comes at 7.06 every morning and takes us to school", because arriving at a reliable frequency count of "at (hour)" in the BNC turned out to be rather tricky even with the help of the Simple Query Syntax Guide for BNCweb available online.

48. "Me and My Girl" is a book _____ love and young people. (about, on)¹⁰¹
49. You look tired. Why don't you lie _____ and relax? (down)
50. Here's another photo of my family. My baby brother is sitting next _____ my father. (to)

Test for intermediate students (B1)
(Gitsaki, 1999)

1. She is a good student. She _____ her homework every day. (does)
2. I haven't finished my essay yet. _____ fact I've only just started it. (in)¹⁰²
3. John is your new boss. Convince him _____ your ability to type fast. (of)
4. He is a very nice person. They all respect him and think _____ of him. (highly)
5. Keep your hands _____ from the cake. It's for tomorrow's party. (away)
6. He went away to university. He has a degree _____ chemistry. (in)
7. The National Gallery is _____ an exhibition of modern art. (holding, having, hosting)
8. I've got skills _____ the use of computers. (in)
9. He is engaged _____ Mary. Their wedding is in two months. (to)
10. The van was full _____ students. They were going to a party. (of)
11. Doctors fail _____ their duty if they do not warn their patients of the dangers. (in)¹⁰³
12. You are a police officer. Find _____ who the thief is. (out)
13. I looked _____ his picture but couldn't recognize him. (at)
14. She played for the tennis team _____ the fact that she was injured. (despite)
15. I play basketball _____ Fridays. (on)
16. Their business is doing very _____. They're making a lot of money. (well)
17. I braked _____ but the car did not stop. (hard)
18. There were fires everywhere. I was _____ a panic. (in)
19. The zoo is crowded _____ visitors every day. (with)
20. Suddenly the lights went _____. I couldn't see anything. (out)
21. Helen just left. She must be _____ her way to school. (on)
22. Marco Polo sailed from Venice _____ the year 1271. (in)
23. He is a good brother. He is looking _____ his little brother. (after)

¹⁰¹ In Sentence 48 the noun "musical" was replaced with the noun "book", as we felt our subjects were likely to be more familiar with the latter.

¹⁰² In Sentence 2 the word "only" was added.

¹⁰³ Sentence 11 ("He failed _____ (in) his university examinations") was replaced with "Doctors fail _____ (in) their duty if they do not warn their patients of the dangers." The original sentence remains grammatically correct when the preposition "in" is omitted.

24. Can you _____ shorthand, Ms Morris? (take)
25. He was _____ the truth but nobody believed him. (telling)
26. You are staying in a hotel. Complain _____ the air-conditioning. It's too cold. (about)
27. I had a nightmare last night. I dreamt that I was _____ trial for murder. (on)
28. He looks terrible. I think he's _____ the penalty for all those late nights. (paying)¹⁰⁴
29. You are getting fat. You should _____ your weight. (watch)
30. He was an unhappy child. He grew _____ without his parents. (up)
31. She always _____ care of her little sister. (takes)
32. I looked _____ my book everywhere and I can't find it. (for)¹⁰⁵
33. Listen to these sentences. Note _____ on a piece of paper the incomplete ones. (down)
34. Did you _____ a call to John? (make)¹⁰⁶
35. Give your friend directions _____ the use of your new computer game. (on)
36. They are Christians. They believe _____ Christ. (in)
37. When she went to Italy, her family came to the airport to see her _____. (off)
38. If you are driving a car you have to stop when the _____ light is red. (traffic)
39. Helen is a student. She is studying English _____ a British university. (at)
40. People will have no hair _____ the future. (in)
41. Sam is not lazy. _____ the contrary, he is a good student. (on)
42. He gave me only one _____ of information that was very important to me. (piece)
43. She was yelling _____ help but nobody could hear her. (for)
44. The radio said that the weather is clearing _____ for the weekend. (up)
45. Find Unit 10 _____ your book. (in)
46. These are two pairs of running shoes. _____ a comparison between the two. (make)
47. You have to work _____ to become a millionaire. (hard)
48. You are painting your room. Ask _____ help from your brother. (for)
49. Your homework for tomorrow is _____ page 86. (on)
50. I'm going to Greece next month. _____ the way, do you remember Eleni, my Greek friend? She is taking me to an island. (by)
51. After you finish your phone call you should hang _____ immediately. (up)
52. This is a photo from my birthday party. The boy _____ my right is my boyfriend. (on)
53. It's 7 o'clock. Time to get _____ of bed and go to school. (out)
54. She was sympathetic _____ my problems. (to)
55. I braked but the car behind me couldn't stop _____ time. (in)

¹⁰⁴ Sentence 28 ("The referee _____ a penalty and the game stopped") was replaced by "He looks terrible. I think he's _____ the penalty for all those late nights." We felt that the original sentence might be a bit problematic for those test-takers who are not too familiar with the rules of football/soccer.

¹⁰⁵ In Sentence 32 "...and I can't find it" was added.

¹⁰⁶ In Sentence 34 an English name John was used.

56. If you have a negative attitude _____ computers, you won't be able to understand them.
(towards, to)
57. As I was driving I was listening to music _____ the radio. (on)
58. It was in the newspaper the other day that mother had found her missing son _____
accident. (by) ¹⁰⁷
59. After you complete the test, add _____ your scores. (up)
60. The war took _____ in Sparta. (place)
61. Before you leave the restaurant you have to fill _____ a questionnaire. (in, out)
62. Mary was very upset _____ her health. (about)
63. The police officer asked them what was going _____ inside. (on)
64. The accusations _____ Peter were very serious, and he was put in prison. (against) ¹⁰⁸
65. He was working every day _____ his painting. He wanted to finish it before Christmas.
(on)

¹⁰⁷ The headline "Mother Finds Missing Son _____ Accident" in Sentence 58 was replaced by "that mother had found her missing son _____ accident." We felt this would be easier for students to understand.

¹⁰⁸ Re Sentence 64 ("The accusations _____ Peter made the judge put him in prison"), the latter part of the sentence ("...made the judge put him in prison") was replaced by "were very serious, and he was put in prison."

Table A.1. The BBI types and BNC frequencies of occurrence of English collocations tested in the A2 Gitsaki Collocation Test

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no
noun + prep (G1)	book about ¹⁰⁹	book on	book about books about book on books on	306 140 725 572	446 1297	48
prep + noun (G4)	at the moment in the evening in the garden on an island		at the moment at that moment at this moment in the evening in (**) ¹¹⁰ garden on a/an/the (*) island	5114 584 425 1197 1693 847	6123 1197 1693 847	26 37 46 27
adj + prep (G5)	famous for good at next to north of		famous for better at best at good at next to north of (other directions) of	450 358 117 1552 2894 2194 4249	450 2027 2894 6443	38 1 50 29
verb + prep (G8-D)	ask about be on (of a programme) help sb with relax with sth take sth out (of a bag)		ask about asked about asking about asks about be/is/are on help (*) with helped (*) with helping (*) with helps (*) with relax with take (*) out takes (*) out taking (*) out took (*) out	270 547 121 23 84 559 207 127 21 45 1427 97 441 1035	961 84 914 45 3000	33 42 40 47 36
phrasal verb (G) ¹¹¹	get off get up		get off getting off gets off got off get up getting up gets up	860 94 48 440 1719 392 168	1442 3876	15 28

¹⁰⁹ The words in bold are the missing items in the test.

¹¹⁰ (*) – one optional word; (**) – two optional words; **sb** – somebody; **sth** – something.

¹¹¹ Phrasal verbs are not part of the BBI typology (Benson, Benson & Ilson, 1986). Benson, Morton, Benson, Evelyn & Ilson, Robert (1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins Publishing Company. Benson, Morton, Benson, Evelyn & Ilson, Robert (1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins Publishing Company. Benson, Morton, Benson, Evelyn & Ilson, Robert (1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins Publishing Company. Benson, Morton, Benson, Evelyn & Ilson, Robert (1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins Publishing Company.

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no
	hold on		got up held on hold on holding on holds on	1597 1641 856 216 27	2740	31
	lie down		lay (*) down lying (*) down lie (*) down lies (*) down	674 233 429 25	1361	49
	watch out		watch out watching out watched out	405 20 2	427	44
	write down		write down writes down writing down written down wrote down	475 17 84 321 98	995	24
verb + noun (L1)	collect stamps		collect (**) stamps collected (**) stamps collecting (**) stamps collects (**) stamps	10 2 8 2	22	3
	make a diagnosis		make a/an/the (*) diagnosis making a/an/the (*) diagnosis made a/an/the (*) diagnosis	14 8 4	26	7
		establish	establish a/an/the (*) diagnosis establishing a/an/the (*) diagnosis	11 2	13	
		give	give a/a/an (*) diagnosis	4	4	
		reach	reach a/an/the (*) diagnosis reaching a/an/the (*) diagnosis reached a/an/the (*) diagnosis	2 1 2	5	
	make a suggestion		make a/an/the (*) suggestion makes a/an/the (*) suggestion making a/an/the (*) suggestion made a/an/the (*) suggestion	45 6 6 25	82	18
	take a trip		take a/an/the (*) trip taken a/an/the (*) trip taking a/an/the (*) trip took a/an/the (*) trip	69 2 18 19	108	6
		make a	make a/an/the (*)	74	166	

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no
verb + noun (L1)	take notes	trip	trip makes a/an/the (*) trip making a/an/the (*) trip made a/an/the (*) trip take (*) notes taken (*) notes takes (*) notes taking (*) notes took (*) notes	5 28 59 73 8 2 69 17	169	14
		make notes	make (*) notes makes (*) notes making (*) notes made (*) notes	117 7 75 40	239	
adj + noun (L3)	sore throat		sore throat	107	107	22
		bad throat	bad throat	8	8	
		dry throat	dry throat	12	12	
noun + verb (L4)	time flies		time flies	3	16	2
			time flew	8		
			time flying	5		
noun + (of) + noun (L5)	school bus		school bus school buses	57 8	65	43
	member of a club		member of a/an/the (**) club	101	101	30
	phone number		phone number phone numbers	284 96	380	20
		telephone number	telephone number telephone numbers	435 155	590	
		mobile number	mobile number piece of (*) advice pieces of (*) advice	2 59 12	2 71	35
	piece of advice					
		bit of advice	bit of (*) advice bits of (*) advice	19 5	24	
		word of advice	word of (*) advice words of (*) advice	27 14	41	
	points of the compass (replaced by "slice of bread")		points of the compass	20	20	9
	post office		slice/s of (*) bread post office post offices	140 1323 94	140 1417	(9) 13

Note: This table does **not** contain the collocations tested in both A2 and B1 tests. These are presented in Table A.3.

Table A.2. The BBI types and BNC frequencies of occurrence of English collocations tested in the B1 Gitsaki Collocation Test

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no
noun + prep (G1)	accusations against ¹¹²	attitude to	accusation (*) ¹¹³ against accusations (*) against accusation/s of (*) (*) against attitude (*) towards attitudes (*) towards attitude (*) to attitudes (*) to degree in degrees in	18 32 6 579 462 1216 712 471 172	56 1041 1928 643	64 56 6
	attitude towards		direction on directions on skill/s in	16 10 738	26 738	35 8
	degree in					
	directions on					
	skills in					
prep + noun (G4)	by accident		by (*) accident	365	365	58
	by the way		by the way	1542	1542	50
	despite the fact		despite the fact	836	836	14
	in a panic		in a/the (*) panic	85	85	18
	in fact		in fact	4790	4790	2
	in a book		in a/the (*) book in my, your (etc.) book in (*)'s book in the future	1633 642 84 2551	1633	45 40
	in the future		in the year in time	1581 3560	1581 3560	22 55
	in the year (1271)					
	in time					
	on page (e.g. 86)		on a/the page on page (number)	269 1391	1660	49
	on the contrary		on the contrary	798	798	41
	on the radio		on the radio	428	428	57
	on trial		on (*) trial/s	448	448	27
adj + prep (G5)	engaged to	sympathetic towards	engaged to full of	31 5535	31 5535	9 10
	full of		sympathetic to	318	318	54
	sympathetic to		sympathetic towards upset about	35 124	35 124	62
	upset about					
verb + prep (G8-D)	believe in		believe (*) in believed (*) in believes (*) in believing (*) in	1576 565 242 115	2498	36
	convince sb of		convince (*) of	79	123	3

¹¹² The words in bold are the missing items in the test.

¹¹³ (*) – one optional word; (**) – two optional words; **sb** – somebody; **sth** – something.

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no
verb + prep (G8-D)	fail in		convinced (*) of convinces (*) of convincing (*) of fail in fails in failing in failed in yell for yelling for yelled for yells for	32 2 10 109 37 82 278 5 10 10 1	506 26	11 43
	yell for					
phrasal verb (G) ¹¹⁴	add up		add (*) up adding (*) up added (*) up adds (*) up	457 109 107 149	822 681	59 44
	clear up		clear up clears up cleared up clearing up	304 32 212 133	 3790	 53
	get out		get out getting out gets out gotten out	3155 485 147 3	 3963	 20
	go out		go out goes out going out gone out went out	2056 238 2 375 1292	 2073	 30
	grow up		grow up grown up grows up grew up	573 686 94 720	 2142	 51
	hang up		hang on hanging on hangs on hanged on hung on	1372 347 58 6 359	 83	 33
	note down		note down noted down notes down noting down	39 28 2 14	 282	 37
	see sb off		see (*) off seeing (*) off saw (*) off seen (*) off	110 36 78 58	 	
verb + noun (L1)	pay the penalty		pay a/an/the (*) penalty pays a/an/the (*) penalty paying a/an/the (*) penalty paid a/an/ the (*) penalty	25 3 7 22	57 48	28 7
	hold an exhibition		held a/an/the (*) exhibition/s hold a/an/the (*) exhibition/s holding a/an/the (*) exhibition/s	17 8 19	 	

¹¹⁴ Phrasal verbs are not part of the BBI typology (Benson, Benson & Ilson, 1986).

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no								
verb + noun (L1)		have an exhibition	holds a/an/the (*) exhibition/s	4	52	34								
			have a/an/the (*) exhibition/s	16										
			has a/an/the (*) exhibition/s	18										
			had a/an/the (*) exhibition/s	18										
			host an exhibition	host a/an/the (*) exhibition/s			7	17						
				hosting a/an/the (*) exhibition/s			9							
				hosted a/an/the (*) exhibition			1							
				make a call			make a/an/the (*) call/s		83	295				
							makes a/an/the (*) call/s		57					
							making a/an/the (*) call/s		70					
							make a comparison		made a/an/the (*) call/s		85	173		
									make a/an/the (*) comparison/s		92			
		making a/an/the (*) comparison/s			50									
		take care			makes a/an/the (*) comparison/s				17		1579			
					made a/an/the (*) comparison/s				14					
					take care				978					
			take place		taking care			219	5000					
					taken care			177						
					takes care			73						
				take shorthand	took care			132		24				
					take place			1538						
					taking place			701						
					tell (sb) the truth		takes place	729				894		
							taken place	715						
							took place	1317						
		watch one's weight					take shorthand	3			18			
							taking shorthand	2						
							took shorthand	2						
							tell (*) the truth	510	29					
							telling (*) the truth	238						
							tells (*) the truth	18						
							told (*) the truth	128		112				
							watch (*) weight	12						
							watching (*) weight	5						
							watches (*) weight	1				112		
														65
													266	
						65								
									266					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					
										42				
											42			
												42		
													42	
								42						
									42					

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no
		informatio n	information			
	traffic light		traffic light/s	308	308	38
adv + adj (L6)	very well		very well	3862	3862	16
verb + adverb (L7)	brake hard		brake hard braked hard braking hard brakes hard	9 9 4 2	24	17
	think highly (of)		think highly of thought highly of thinks highly of	7 5 2	14	4

Note: This table does **not** contain the collocations tested in both A2 and B1 tests. These are presented in Table A.3.

Table A.3. The BBI types and BNC frequencies of occurrence of English collocations tested in both A2 and B1 Gitsaki Collocation Tests

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no (A2 test)	Item no (B1 test)
prep + noun (G4)	at ¹¹⁵ a university		at a/an/the university at a/an/the (*) ¹¹⁶ university	1151 75	1226	17	39
	on sb's right		on my, your (etc.) right on the right on (*)'s right on (*) (*)'s right	369 1451 8	1820	8	52
	on one's way		on my, your (etc.) way on the way on (*)'s way	12 2886 2166 1	5053	23	21
	on Friday/Sund ay		on Friday/s on Sunday/s on (other days of the week)	1815 2352 9778	13945	19	15
adj + prep (G5)	crowded with		crowded with	170	170	5	19
verb + prep (G8-D)	ask for		ask (*) for asked (*) for asking (*) for	1825 1839 1102	4937	25	48

¹¹⁵ The words in bold are the missing items in each test.

¹¹⁶ (*) – one optional word; (**) – two optional words; **sb** – somebody; **sth** – something.

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no (A2 test)	Item no (B1 test)	
verb + prep (G8-D)	complain about		asks (*) for	171	827	21	26	
			complain (*) about	325				
			complained (*) about	256				
			complaining (*) about	215				
			complains (*) about	31				
	look at		look (*) at	4988	16090	12	13	
			looking (*) at	4991				
			looked (*) at	5013				
	look for		looks (*) at	1098	5978	34	32	
			look for	2227				
			looked for	293				
	work on sth		looking for	3358	5983	39	65	
			looks for	100				
			work (*) on	1665				
			working (*) on	2795				
			works (*) on	581				
			worked (*) on	942				
phrasal verb	fill in	fill out	fill in	782	1482	16	61	
			filling in	252				
			filled in	412				
			fills in	36				
			fill out	122				
			filling out	39				
	find out		filled out	63	5000	10	12	
			fills out	10				
			find out	3558				
			finding out	413				
			finds out	90				
			found out	939				
	go on		go on	1716	4820	45	63	
			goes on	467				
			going on	1269				
	keep sb/sth away		went on	1368	522	32	5	
			keep (**) away	342				
			keeping (**) away	42				
			kept (**) away	138				
	look after		look after	2233	4360	11	23	
			looking after	1072				
			looked after	848				
			looks after	207				
	verb + noun (L1)		do one's homework		did (*)		174	4
homework do (*)					17			
homework does (*)					69			
homework doing (*)					4			
					37			

Type of phrase (G - grammatical, L - lexical)	Base forms	Other possible answers	Inflected forms	Frequencies of inflected forms	Overall frequency	Item no (A2 test)	Item no (B1 test)
			homework done (*) homework	47			
verb + adverb (L7)	work hard		work (*) hard worked (*) hard working (*) hard works (*) hard	511 535 301 68	1415	41	47

APPENDIX F: Vocabulary Levels Tests

1,000-word level

Instructions:¹¹⁷

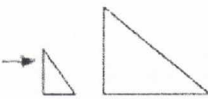
Write **T** if a sentence is **true**. Write **F** if it is **not true**. Write **X** if you do not understand the sentence.

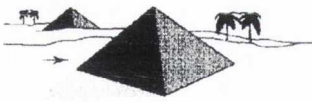
- T – true
- F – false
- X – do not understand the sentence

Here is an example.

We can stop time. F
We divide time into minutes, hours and days¹¹⁸. T

Version 1 (Nation, 2001b)

1. This one is little.  (True)

2. You can find these everywhere.  (False)

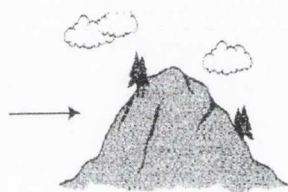
- 3. Some children call their mother Mama. (True)
- 4. “Show me the way to do it” means “show me how to do it”. (True)
- 5. This country is part of the world. (True)

- 6. This can keep people away from your house.  (True)
- 7. When something falls, it goes up. (False)
- 8. Most children go to school at night. (False)
- 9. It is easy for children to remain still. (False)

¹¹⁷ Instructions were provided in Polish and English.
¹¹⁸ “Cut time” was used instead of “divide time” in the original version of the instructions.



10. One person can carry this. (False)
11. A scene is a part of a play. (True)
12. People often think of their home when they are away from it. (True)
13. There is a mountain in every city. (False)
14. Each month has the same number of days. (False)
15. A chief is the youngest person in a group. (False)
16. Black is a colour. (True)
17. You can use a pen to make marks on paper. (True)
18. A family always has at least two people. (True)
19. You can go by road from London to New York. (False)
20. Gold¹¹⁹ costs a lot of money. (True)



21. This is a hill. (True)



22. This young person is a girl.¹²⁰ (True)
23. We can be sure that one day we will die. (True)
24. A society is made of people living together. (True)
25. An example can help you understand. (True)
26. Some books have pictures in them. (True)
27. When some people attack other people, they try to hurt them. (True)
28. When something is ancient, it is very big. (False)
29. Big ships can sail up a stream. (False)
30. It is good to keep a promise. (True)
31. People often dream when they are sleeping. (True)
32. This is a date – 10 o'clock. (False)

¹¹⁹ "Silver" from the original version of the test was replaced by "gold" (both within the 2,000-word frequency band). Silver is regarded as relatively cheap in Poland, while in Ireland it appears to be on the expensive side. Gold, on the other hand, is perceived as expensive in both countries.

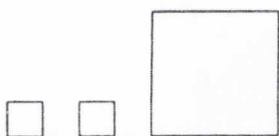
¹²⁰ The original picture of the girl was replaced by another one, as it could easily be taken for an image of a boy with longer hair.

33. When something is impossible, it is easy to do. (False)
34. Milk is blue. (False)
35. A square has five sides. (False)
36. Boats are made to travel on land. (False)
37. Cars cannot pass each other on a wide road. (False)
38. When you look at something closely, you can see the details. (True)

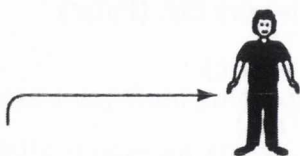


39. This part is a handle. (True)

Version 2 (Nation, 2001b)

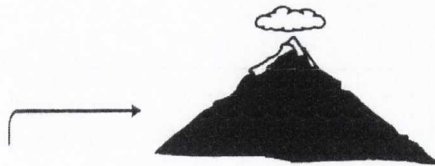


1. Two of these are little. (True)
2. You must look, when you want to find the way. (True)
3. When someone asks 'What is your dog called?' you should say your dog's name.¹²¹ (True)
4. There are many ways to get money. (True)
5. All the world is under water. (False)
6. When you keep asking, you ask once. (False)
7. Sometimes people die when they fall off a building. (True)
8. Day follows night and night follows day. (True)
9. "Remain here" means "stay". (True)

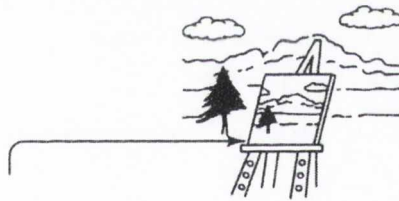


10. This is a person. (True)
11. When there is a change of scene, we see a different place. (True)
12. "Often" means "many times". (True)

¹²¹ The original sentence was as follows: "When someone asks 'What are you called?' you should say your name."



13. This is a mountain. (True)
14. Each month has a different name. (True)
15. People follow the orders¹²² of a chief. (True)
16. Green is a colour. (True)
17. Dirty hands cannot leave marks on glass. (False)
18. You need at least five people to make a group. (False)
19. Cars move along¹²³ a road. (True)
20. You can eat silver. (False)
21. You can see more when you are on a hill. (True)
22. Your child will be a girl or a boy. (True)
23. When you are sure, you know you are right. (True)
24. Each society has the same rules. (False)
25. Three examples of food are: shops, homes and markets. (False)



26. This is a picture. (True)
27. It is good to attack people. (False)
28. Rome (Rzym) is an ancient city. (True)¹²⁴
29. A stream is a small river. (True)
30. When you promise something, you say you will really do it. (True)
31. Dreams are about things that really happened. (False)
32. When we give a date, we say the day, the month and the year. (True)
33. It is impossible to live for a long time without water. (True)
34. Very young children drink milk. (True)

¹²² In the original version "follow orders" was used (no article).

¹²³ "On" was used instead of "along" in the original version of the test.

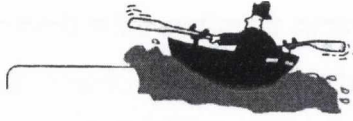
¹²⁴ The Polish name for Rome ("Rzym") was added in brackets, as it was feared that not every student may be familiar with the English name. Two other city names used in this test, New York and London, are so widely known that we did not feel the need to provide the Polish equivalents.

35. This is a square.



(True)

36. This is a boat.



(True)

37. It is a short way from one side to the other side of a wide river. (False)

38. A detail is a small piece of information. (True)

39. A handle is part of our body. (False)

Vocabulary Levels Test:

2,000 and 3,000-word level

Instructions:

This is a vocabulary test. You must choose the right word (A-F) to go with each meaning (1-30). Here is an example.

- | | |
|-------------|-------------------------------|
| A. business | 1. part of a house |
| B. clock | 2. animal with four legs |
| C. horse | 3. something used for writing |
| D. pencil | |
| E. shoe | |
| F. wall | |

You answer it in the following way.

1. F
2. C
3. D

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are "business", "clock" and "shoe".

If you have no idea about the meaning of a word, do not guess. If you think you might know the meaning, then you should try to find the answer.

Version 1
(Schmitt, 2000: 192-195)

The 2,000-word level

- | | |
|----------------|---|
| A. birth | 4. game (E) |
| B. dust | 5. winning (F) |
| C. operation | 6. being born (A) |
| D. row | |
| E. sport | |
| F. victory | |
| | |
| A. choice | 7. heat (F) |
| B. crop | 8. meat (C) |
| C. flesh | 9. money paid regularly for doing a job (D) |
| D. salary | |
| E. secret | |
| F. temperature | |
| | |
| A. cap | 10. teaching and learning (B) |
| B. education | 11. numbers to measure with (E) |
| C. journey | 12. going to a far place (C) |
| D. parent | |
| E. scale | |
| F. trick | |
| | |
| A. attack | 13. gold and silver (F) |
| B. charm | 14. pleasing quality (B) |
| C. lack | 15. not having something (C) |
| D. pen | |
| E. shadow | |
| F. treasure | |
| | |
| A. cream | 16. part of milk (A) |
| B. factory | 17. a lot of money (F) |
| C. nail | 18. person who is studying (D) |
| D. pupil | |
| E. sacrifice | |
| F. wealth | |
| | |
| A. adopt | 19. go up (B) |
| B. climb | 20. look at closely (C) |
| C. examine | 21. be on every side (F) |
| D. pour | |
| E. satisfy | |
| F. surround | |
| | |
| G. bake | 22. join together (B) |
| H. connect | 23. walk without purpose (F) |
| I. inquire | 24. keep within a certain size (D) |
| J. limit | |
| K. recognize | |
| L. wander | |

- A. burst
- B. concern
- C. deliver
- D. fold
- E. improve
- F. urge

- A. original
- B. private
- C. royal
- D. slow
- E. sorry
- F. total

- A. brave
- B. electric
- C. firm
- D. hungry
- E. local
- F. usual

- 25. break open (A)
- 26. make better (E)
- 27. take something to someone (C)

- 28. first (A)
- 29. not public (B)
- 30. all added together (F)

- 31. commonly done (F)
- 32. wanting food (D)
- 33. having no fear (A)

The 3,000-word level

- A. belt
- B. climate
- C. executive
- D. notion
- E. palm
- F. victim

- A. acid
- B. bishop
- C. chill
- D. ox
- E. ridge
- F. structure

- A. bench
- B. charity
- C. jar
- D. mate
- E. mirror
- F. province

- A. boot
- B. device
- C. lieutenant
- D. marble
- E. phrase
- F. vein

- 1. idea (D)
- 2. inner surface of your hand (E)
- 3. strip of leather worn around the waist (A)

- 4. cold feeling (C)
- 5. farm animal (D)
- 6. organization or framework (F)

- 7. long seat (A)
- 8. help to the poor (B)
- 9. part of a country (F)

- 10. army officer (C)
- 11. a kind of stone (D)
- 12. tube through which blood flows (F)

- A. candle
- B. draft
- C. horror
- D. prospect
- E. residence¹²⁵
- F. timber

- 13. a state of living in a place (E)
- 14. chance of something happening (D)
- 15. first rough form of something written (B)

- A. betray
- B. dispose
- C. embrace
- D. injure
- E. proclaim
- F. scare

- 16. frighten (F)
- 17. say publicly (E)
- 18. hurt seriously (D)

- A. encounter
- B. illustrate
- C. inspire
- D. plead
- E. seal
- F. shift

- 19. meet (A)
- 20. beg for help (D)
- 21. close completely (E)

- A. assist
- B. bother
- C. condemn
- D. erect
- E. trim
- F. whirl

- 22. help (A)
- 23. cut neatly (E)
- 24. spin around quickly (F)

- A. annual
- B. concealed
- C. definite
- D. mental
- E. previous
- F. savage

- 25. wild (F)
- 26. clear and certain (C)
- 27. happening once a year (A)

- A. dim
- B. junior
- C. magnificent
- D. maternal
- E. odd
- F. weary

- 28. strange (E)
- 29. wonderful (C)
- 30. not clearly lit (A)

Version 2

(Nation, 2001b) modified by Schmitt et al. (2001: 82-84)

The 2,000-word level

- A. copy
- B. engine¹²⁶
- C. event
- D. pity
- E. profit
- F. tip

- 1. end or highest point (F)
- 2. this moves a car (B)
- 3. thing made to be like another (A)

¹²⁵ The word “apartment” was replaced by the word “residence” (see Section 5.2.2 for more information).

¹²⁶ The word “motor” was replaced by the word “engine” (Section 5.2.2).

- | | |
|----------------|--|
| A. accident | 4. loud deep sound (E) |
| B. debt | 5. something you must pay (B) |
| C. fortune | 6. having a high opinion of yourself (D) |
| D. pride | |
| E. roar | |
| F. thread | |
| | |
| A. coffee | 7. money for work (F) ¹²⁷ |
| B. disease | 8. a piece of clothing (D) |
| C. justice | 9. using the law in the right way (C) |
| D. skirt | |
| E. stage | |
| F. wage | |
| | |
| A. clerk | 10. a drink (F) |
| B. frame | 11. office worker (A) |
| C. noise | 12. unwanted sound (C) |
| D. respect | |
| E. theatre | |
| F. wine | |
| | |
| A. dozen | 13. chance (D) |
| B. empire | 14. twelve (A) |
| C. gift | 15. money paid to the government (F) |
| D. opportunity | |
| E. relief | |
| F. tax | |
| | |
| A. admire | 16. make wider or longer (F) |
| B. complain | 17. bring in for the first time (E) |
| C. fix | 18. have a high opinion of someone (A) |
| D. hire | |
| E. introduce | |
| F. stretch | |
| | |
| A. arrange | 19. grow (B) |
| B. develop | 20. put in order (A) |
| C. lean | 21. like more than something else (E) |
| D. owe | |
| E. prefer | |
| F. seize | |
| | |
| A. blame | 22. make (D) |
| B. elect | 23. choose by voting (B) |
| C. jump | 24. become like water (E) |
| D. manufacture | |
| E. melt | |
| F. threaten | |

¹²⁷ The cluster “game” (“sport”/ E), “winning” (“victory”/ F) and “being born” (“birth”/ A) was removed by Schmitt et al. (2001) from Version 1, as this very same cluster is repeated in Version 2 of this test.

- A. ancient
- B. curious
- C. difficult
- D. entire
- E. holy
- F. social

- 25. not easy (C)¹²⁸
- 26. very old (A)
- 27. related to God (E)

- A. bitter
- B. independent
- C. lovely
- D. merry
- E. popular
- F. slight

- 28. beautiful (C)
- 29. small (F)
- 30. liked by many people (E)

The 3,000-word level

- A. bull
- B. champion
- C. dignity
- D. hell
- E. museum
- F. solution

- 1. formal and serious manner (C)
- 2. winner of a sporting event (B)
- 3. building where valuable objects are shown (E)

- A. blanket
- B. contest
- C. generalization
- D. merit
- E. plot
- F. vacation

- 4. holiday (F)
- 5. good quality (D)
- 6. wool covering used on beds (A)

- A. comment
- B. gown
- C. import
- D. nerve
- E. pasture
- F. tradition

- 7. long formal dress (B)¹²⁹
- 8. goods from a foreign country (C)
- 9. part of the body which carries feeling (D)

- A. administration
- B. angel
- C. frost
- D. herd
- E. fort
- F. pond

- 10. group of animals (D)
- 11. spirit who serves God (B)
- 12. managing business and affairs (A)

- A. atmosphere
- B. counsel
- C. factor
- D. hen
- E. lawn
- F. muscle

- 13. advice (B)
- 14. a place covered with grass (E)
- 15. female chicken (D)

¹²⁸ The cluster “commonly done” (“usual”/ F), “wanting food” (“hungry”/ D) and “having no fear” (“brave”/ A) was removed by Schmitt et al. (2001) from Version 1, as this very same cluster is repeated in Version 2 of this test.

¹²⁹ The cluster “a place to live” (“apartment”/ A), “chance of something happening” (“prospect”/ E) and “first rough form of something written” (“draft”/ C) was removed by Schmitt et al. (2001) from Version 1 for the same reason as above.

- A. abandon
- B. dwell
- C. oblige
- D. pursue
- E. quote
- F. resolve

- 16. live in a place (B)
- 17. follow in order to catch (D)
- 18. leave something permanently (A)

- A. assemble
- B. attach
- C. peer
- D. quit
- E. scream
- F. toss

- 19. look closely (C)
- 20. stop doing something (D)
- 21. cry out loudly in fear (E)

- A. drift
- B. endure
- C. grasp
- D. knit
- E. register
- F. tumble

- 22. suffer patiently (B)
- 23. join wool threads together (D)
- 24. hold firmly with your hands (C)

- A. brilliant
- B. distinct
- C. magic
- D. naked
- E. slender
- F. stable

- 25. thin (E)
- 26. steady (F)
- 27. without clothes (D)

- A. aware
- B. blank
- C. desperate
- D. normal
- E. striking
- F. supreme

- 28. usual (D)
- 29. best or most important (F)
- 30. knowing what is happening (A)

Productive test
(Skrzypek, developed in 2007)

The 1,000-word level

1. He is an ac_____ working in a theatre. (actor)
2. His first ch_____, a boy called Peter, died at three years of age. (child)
3. I don't believe a word she says. She's such a li_____! (liar)
4. I'm afraid I don't have much mo_____ on me – only 5 Euro or so. (money)
5. I get up quite early in the mo_____ – at 6.30 a.m. (morning)
6. I live near Patty now and there's this great feeling of having a really good nei_____.
(neighbour)
7. Those pink cl_____ in the sky looked strange. (clouds)
8. He doesn't know how to behave at the table. Somebody needs to teach him good
ma_____. (manners)

9. In the winter when there is a bit of sn____, we'll go skiing in the mountains. (snow)
10. When you go to a restaurant, what do you normally or____? (order)
11. I am finding it difficult to p____ all the bills. Everything is so expensive! (pay)
12. What's her name? I can never rem____ it. (remember)
13. Read the questions carefully and cir____ the correct answer. (circle)
14. I fi____ work quite late and then I head back home. (finish)
15. The story was very funny but Mr Brown did not sm____. (smile)
16. Thank you very much for your help. That was very k____ of you. (kind)
17. I can't carry all those he____ things. They're 5 kilos each! (heavy)
18. Yes, I am mad about my boyfriend, but I am too young to get ma____. (married)

Productive test: Version 1
(Laufer & Nation, 1999: 46-47)

The 2,000-word level

1. I'm glad we had this opp____ to talk. (opportunity)
2. There are a **doz**¹³⁰ ____ eggs in the basket. (dozen)
3. Every working person must pay income t____. (tax)
4. The pirates buried the trea____ on a desert island. (treasure)
5. Her beauty and cha____ had a powerful effect on men. (charm)
6. La____ of rain led to a shortage of water in the city. (lack)
7. He takes cr____ and sugar in his coffee. (cream)
8. The rich man died and left all his we____ to his son. (wealth)
9. **Pup**____ must hand in their papers by the end of the week. (pupils)
10. This sweater is too tight. It needs to be **stret**____. (stretched)
11. Ann **intro**____ her boyfriend to her mother. (introduced)
12. Teenagers often **adm**____ and worship pop singers. (admire)
13. If you blow up that balloon any more it will **bur**____. (burst)
14. In order to be accepted into the university, he had to **impr**____ his grades. (improve)
15. The telegram was **deli**____ two hours after it had been sent. (delivered)
16. The differences were so sl____ that they went unnoticed. (slight)
17. The dress you're wearing is lov____. (lovely)
18. He wasn't very **popu**____ when he was a teenager, but he has many friends now.
(popular)

¹³⁰ The prompts in bold provide most of the word stem.

The 3,000-word level

1. He has a successful car_____ as a lawyer. (career)
2. The thieves threw ac_____ in his face and made him blind. (acid)
3. To improve the country's economy, the government decided on economic ref_____.
(reform)
4. She wore a beautiful green go_____ to the ball. (gown)
5. The government tried to protect the country's industry by reducing the imp_____ of
cheap goods. (import)
6. The children's games were funny at first, but finally got on the parents' **ner**¹³¹_____.
(nerves)
7. The lawyer gave some wise **coun**_____ to his client. (counsel)
8. Many people in England mow the la_____ of their houses on Sunday morning. (lawn)
9. The farmer sells the eggs that his **he**_____ lays. (hen)
10. Sudden noises at night **sca**_____ me a lot. (scare)
11. France was proc_____ a republic in the 18th century. (proclaimed)
12. Many people are inj_____ in road accidents every year. (injured)
13. Suddenly he was **thru**_____ into the dark room. (thrust)
14. He perc_____ a light at the end of the tunnel. (perceived)
15. Children are not independent. They are att_____ to their parents. (attached)
16. She showed off her sle_____ figure in a long narrow dress. (slender)
17. She has been changing partners often because she cannot have a sta_____ relationship
with one person. (stable)
18. You must wear a bathing suit on a public beach. You're not allowed to be na_____.
(naked)

¹³¹ The prompts in bold provide most of the word stem.

Productive test: Version 2

The 2,000-word level (Laufer & Nation, 1999: 48-49)

1. It is the de_____ that counts, not the thought. (deed)
2. Plants receive water from the soil through their ro_____. (roots)
3. The nu_____ was helping the doctor in the operating room. (nurse)
4. Since he is unskilled, he earns low wa_____. (wages)
5. This year long sk_____ are fashionable again. (skirts)
6. Laws are based on the principle of jus_____. (justice)
7. He is walking on the ti_____ of his toes. (tips)
8. The mechanic had to replace the en_____ of the car. (engine)¹³²
9. There is a co_____ of the original report in the file. (copy)
10. They had to cl_____ a steep mountain to reach the cabin. (climb)
11. The doctor ex_____ the patient thoroughly. (examined)
12. The house was su_____ by a big garden. (surrounded)
13. The railway con_____ London with its suburbs. (connects)
14. She wan_____ aimlessly in the streets. (wandered)
15. The organisers li_____ the number of participants to fifty. (limited)
16. This work is not up to your **usu**¹³³ _____ standard. (usual)
17. They sat down to eat even though they were not hu_____. (hungry)
18. You must have been very br_____ to participate in such a dangerous operation. (brave)

The 3,000-word level (Laufer & Nation, 1999:49)

1. The ambassador's official res_____ is situated in West Street. (residence)¹³⁴
2. The pro_____ of failing the test scared him. (prospect)
3. Before writing the final version, the student wrote several **dra**_____. (drafts)
4. It was a cold day. There was a ch_____ in the air. (chill)
5. The cart is pulled by an o_____. (ox)
6. Anthropologists study the **struc**_____ of ancient societies. (structure)
7. After two years in the Army, he received the rank of lieu_____. (lieutenant)
8. The statue is made of mar_____. (marble)

¹³² The word "motor" was replaced by the word "engine" (see Section 5.2.2).

¹³³ The prompts in bold provide most of the word stem.

¹³⁴ The sentence "I live in a small apa_____ (apartment) on the second floor" was removed from the Levels Test (see Section 5.2.2).

9. Some aristocrats believed that blue blood flowed through their ve _____. (veins)
10. The secretary **assi** _____ the boss in organizing the course. (assisted)
11. His beard was too long. He decided to tr _____ it. (trim)
12. People were **whir** _____ around on the dance floor. (whirling)
13. He was on his knees, **ple** _____ for mercy. (pleading)
14. You'll sn _____ that branch if you bend it too far. (snap)
15. I won't tell anybody. My lips are **sea** _____. (sealed)
16. Crying is a nor _____ response to pain. (normal)
17. The Emperor of China was the **supr** _____ ruler of his country. (supreme)
18. You must be **awa** _____ that very few jobs are available. (aware)

Table A.4. Words tested receptively and productively in Versions 1 & 2 of 2,000 and 3,000 Vocabulary Levels Tests

Level	Target words	Item number			
		Receptive 1 ¹³⁵	Receptive 2	Productive 1	Productive 2
2,000 level tests	admire		18	12	
	brave	30	27		18
	burst	22		13	
	charm	11		5	
	climb	16			10
	connect	19			13
	copy		3		9
	cream	13		7	
	deliver	24		15	
	dozen		14	2	
	engine ¹³⁶		2		8
	examine	17			11
	hungry	29	26		17
	improve	23		14	
	introduce		17	11	
	justice		9		6
	lack	12		6	
	limit	21			15
	lovely		28	17	
	opportunity		13	1	
	popular		30	18	
	pupil	15		9	
	skirt		8		5
	slight		29	16	
	stretch		16	10	
	surround	18			12
	tax		15	3	
	tip		1		7
	treasure	10		4	
	usual	28	25		16
	wage		7		4

¹³⁵ Receptive versions 1 and 2 were devised by Schmitt (2000) & Nation (2001) respectively. Receptive Version 2 was further modified by Schmitt et al. (2001). Productive versions 1 and 2 were devised by Laufer & Nation (1999).
¹³⁶ The word “motor” was replaced by the word “engine” (see Section 5.2.2).

Level	Target words	Item number			
		Receptive 1 ¹³⁵	Receptive 2	Productive 1	Productive 2
3,000 level tests	wander	20			14
	wealth	14		8	
	assist	22			10
	aware		30		18
	chill	4			4
	counsel		13	7	
	draft	15	9		3
	gown		7	4	
	hen		15	9	
	import		8	5	
	injure	18		12	
	lawn		14	8	
	lieutenant	10			7
	marble	11			8
	naked		27	18	
	nerve		9	6	
	normal		28		16
	ox	5			5
	plead	20			13
	proclaim	17		11	
	prospect	14	8		2
	residence ¹³⁷	13	7		1
	scare	16		10	
	seal	21			15
	slender		25	16	
	stable		26	17	
	structure	6			6
	supreme		29		17
	trim	23			11
	vein	12			9
	whirl	24			12

Table A.5. Words tested only receptively in Version 1 & 2 of 2,000 and 3,000 Vocabulary Levels Tests

2,000 level			3,000 level		
Target words	Item number		Target words	Item number	
	Receptive 1 ¹³⁸	Receptive 2		Receptive 1	Receptive 2
ancient		26	abandon		18
arrange		20	administration		12
birth	3		angel		11
clerk		11	annual	27	
debt		5	belt	3	
difficult		25	bench	7	
education	7		blanket		6
elect		23	champion		2
flesh	5		charity	8	
grow		19	definite	26	
holy		27	dignity		1
journey	9		dim	30	
manufacture		22	dwell		16
melt		24	encounter	19	
noise		12	endure		22
original	25		grasp		24
prefer		21	herd		10
pride		6	knit		23
private	26		magnificent	29	

¹³⁷ The word “apartment” was replaced by the word “residence” (see Section 5.2.2).

¹³⁸ Receptive Versions 1 and 2 were devised by Schmitt (2000) & Nation (2001) respectively. Receptive Version 2 (Nation, 2001) at the 2,000-word level was further modified by Schmitt et al. (2001).

2,000 level		Item number		3,000 level		Item number	
Target words		Receptive 1 ¹³⁸	Receptive 2	Target words		Receptive 1	Receptive 2
roar			4	merit			5
salary	6			museum			3
scales	8			notion	1		
sport	1			odd	28		
temperature	4			palm	2		
total	27			peer			19
victory	2			province	9		
wine			10	pursue			17
				quit			20
				savage	25		
				scream			21
				vacation			4

Table A.6. Words tested only productively in Version 1 & 2 of 2,000 and 3,000 Vocabulary Levels Tests

2,000 level		Item number		3,000 level		Item number	
Target words		Productive 2 ¹³⁹		Target words		Productive 1	Productive 2
deed	1			acid	2		
nurse	3			attached	15		
roots	2			career	1		
				perceived	14		
				reform	3		
				snap			14
				thrust	13		

¹³⁹ Productive Versions 1 and 2 were devised by Laufer & Nation (1999).

APPENDIX G: Progress Vocabulary and Collocation Test

Instructions:

This is a vocabulary test. You must choose the right word (A-F) to go with each meaning (1-30). Here is an example.

- | | |
|-------------|-------------------------------|
| A. business | 4. part of a house |
| B. clock | 5. animal with four legs |
| C. horse | 6. something used for writing |
| D. pencil | |
| E. shoe | |
| F. wall | |

You answer it in the following way.

4. F
5. C
6. D
-

A2 receptive progress vocabulary test

- | | |
|----------------|--|
| A. expensive | 1. remarkably good, making you feel happy and satisfied (B) |
| B. wonderful | 2. costing a lot of money (A) |
| C. challenging | 3. done many times in the same way (E) |
| D. helpful | |
| E. repetitive | |
| F. cheap | |
| A. squash | 4. running slowly as a way of exercising (C) |
| B. surfing | 5. travelling by boat (D) |
| C. jogging | 6. making pictures with a pencil or colour markers (E) |
| D. sailing | |
| E. drawing | |
| F. volleyball | |
| A. chef | 7. someone whose job is to arrange funerals (D) |
| B. architect | 8. someone whose job is to examine income, to give tax advice, etc. (E) |
| C. nurse | 9. a skilled cook, especially the main cook in a hotel or restaurant (A) |
| D. undertaker | |
| E. accountant | |
| F. doctor | |
| A. stamp | 10. a network for train travel (D) |
| B. newspaper | 11. the fastest speed, longest distance, etc. especially in sport (F) |
| C. handbag | 12. an occasion when you eat food, for example breakfast or dinner (E) |
| D. railway | |
| E. meal | |
| F. record | |

- | | |
|--------------|---|
| A. forget | 13. like, get pleasure from something (C) |
| B. begin | 14. not to remember facts, information, or people, etc. (A) |
| C. enjoy | 15. start doing something (B) |
| D. suggest | |
| E. hope | |
| F. try | |
| A. collect | 16. say 'no', say that you will not do something that someone has asked you to do (C) |
| B. offer | 17. think that something will happen (F) |
| C. refuse | 18. to get things of the same type from different places and bring them together (A) |
| D. hate | |
| E. continue | |
| F. expect | |
| A. early | 19. behaving towards someone in a way that shows you like them (D) |
| B. glamorous | 20. not interesting in any way (E) |
| C. hot | 21. attractive and successful (B) |
| D. friendly | |
| E. boring | |
| F. hungry | |
| A. outdoor | 22. not feeling well (E) |
| B. busy | 23. without any other people around you (C) |
| C. alone | 24. not inside a building (A) |
| D. different | |
| E. sick | |
| F. exciting | |
| A. bike | 25. money that you receive as payment for your work (F) |
| B. shower | 26. a room where people can work at desks (E) |
| C. question | 27. a piece of equipment that you stand under to wash your whole body (B) |
| D. world | |
| E. office | |
| F. salary | |
| A. company | 28. the person who is in charge of your work (E) |
| B. travel | 29. a group of people who work or play together (C) |
| C. team | 30. the activity of going from one place to another (B) |
| D. routine | |
| E. boss | |
| F. lawyer | |

B1 receptive progress vocabulary test

- | | |
|----------------|---|
| A. mean | 1. the quality of being able to change easily to suit any new situation (B) |
| B. flexible | 2. sure that deal with situations successfully (D) |
| C. charming | 3. willing to give money in order to help people (E) |
| D. confident | |
| E. generous | |
| F. extravagant | |

- A. squash
- B. circuit training
- C. jogging
- D. skateboarding
- E. wrestling
- F. archery

- A. colleague
- B. mystery
- C. chef
- D. charm
- E. trickster
- F. acquaintance

- A. tabloid
- B. supplement
- C. proposal
- D. expansion
- E. character
- F. wealth

- A. return
- B. decrease
- C. arrange
- D. urge
- E. rise
- F. attempt

- A. blame
- B. congratulate
- C. insist
- D. approve
- E. dislike
- F. fancy

- A. dependable
- B. casual
- C. striking
- D. desperate
- E. mutual
- F. opposite

- A. confident
- B. embarrassed
- C. homesick
- D. affordable
- E. up-to-date
- F. spacious

- A. gardening
- B. article
- C. affair
- D. weakness
- E. cover
- F. attachment

- 4. the sport in which two people make each other fall to the ground (E)
- 5. a sport of shooting arrows from a bow (F)
- 6. the activity of running slowly and steadily as a way of exercising (C)
- 7. a special quality someone or something has that makes people like them (D)
- 8. someone who deceives or cheats people (E)
- 9. someone you work with (A)
- 10. a large amount of money or property that a person owns (F)
- 11. a newspaper that has small pages, a lot of photographs, and stories mainly about famous people, etc. (A)
- 12. a separate part of a newspaper or magazine (B)
- 13. increase in number, amount, or value (E)
- 14. strongly suggest that someone does something (D)
- 15. try to do something, especially something difficult (F)
- 16. like or want something (F)
- 17. think that someone is good, right, or suitable (D)
- 18. say or think that someone is responsible for something bad (A)
- 19. unusual and easily noticeable (C)
- 20. as different as possible from something else (F)
- 21. a person who has this quality can be trusted (A)
- 22. feeling nervous and uncomfortable (E)
- 23. modern or fashionable (E)
- 24. reasonably priced (D)
- 25. a fault in someone's character (D)
- 26. a piece of writing about a particular subject in a newspaper or magazine (B)
- 27. a sexual relationship between two people, in which one of them may be married to someone else (C)

- | | |
|---------------|---|
| A. editor | 28. an owner of a business (D) |
| B. columnist | 29. the person who is in charge of a newspaper or magazine (A) |
| C. presenter | 30. someone who writes articles about a particular subject that appear regularly in a newspaper or magazine (B) |
| D. proprietor | |
| E. organizer | |
| F. reader | |

A2 productive progress vocabulary test

- I'll ask my b_____ if I can leave work earlier today. (boss)
- He took a job as a c_____ in a Dublin hotel. He really loves cooking. (chef)
- You should go j_____ every morning. You don't exercise enough. (jogging)
- Lunch is his main m_____ of the day. (meal)
- The local tourist o_____ is located in Tara Street. (office)
- S_____ is one of my hobbies. I've always loved boats and the sea. (sailing)
- In our bathroom we have a bath and a s_____. (shower)
- The football t_____ is not playing very well this season. (team)
- This job involves a lot of air t_____. You'll have to travel by plane very often. (travel)
- The school b_____ in September and ends in June. (begins)
- Her room is full of old baseball bats and hockey sticks. She c_____ old sports equipment. (collect)
- I e_____ playing tennis. I do it every day. (enjoy)
- We are e_____ him to arrive shortly. He should be here any minute now. (expecting)
- What's her name? I always f_____. (forget)
- I offered to help her but she r_____. She didn't want anyone's help. (refuse)
- She was sitting all a_____ in the hall. All the other children were out in the garden. (alone)
- This watch you're wearing looks very e_____. Is it silver? (expensive)
- My job is very boring and r_____. I do the same thing every day. (repetitive)

B1 productive progress vocabulary test

- It's a story of a secret a_____ between a married teacher and a teenage student. (affair)
- My boyfriend wants to take up a_____. He's bought a bow and shooting arrows. (archery)
- She is a c_____ of mine. We work in the same department. (colleague)

4. The e_____ of "The Times" refused to publish information from an unconfirmed source. (editor)
5. He is now the sole p_____ of the business. He's bought out his partner. (proprietor)
6. Vitamin s_____ are recommended for those suffering from vitamin deficiency. (supplements)
7. You mustn't believe in every word they write in the t_____. Most of that is idle gossip. (tabloids)
8. He knows how to fool people. He's such a t_____! (trickster)
9. The sudden w_____ in her legs made her stumble. (weakness)
10. I am very happy to say that the committee a_____ our request for money. We will get the funding soon. (approved)
11. The prisoners a_____ an escape, but failed. (attempted)
12. I am not the one to b_____. It is not my fault. (blame)
13. She didn't f_____ the idea of going out in the dark, and she stayed home. (fancy)
14. The sun r_____ in the east and sets in the west. (rises)
15. The report u_____ that all children be taught to swim. (urged)
16. I find her totally d_____. She can be relied on in any situation. (dependable)
17. I wanted to pay, and then I realised I left my wallet behind. I never felt so e_____ in my life! (embarrassed)
18. She was undoubtedly a very s_____ young woman. Everyone noticed her when she came into the room. (striking)

A2 progress collocation test

Task 1 Multiple-choice questions

Instructions:

This is a collocation test. Please tick (✓) the correct phrase. Here is an example.

- A. do exercise
- B. make exercise

You answer it in the following way.

- A. do exercise ✓
 - B. make exercise
-

- 1. a square of soap
a bar of soap ✓
- 2. play in chess
play chess ✓
- 3. leave the building ✓
leave from the building
- 4. go on the party
go to the party ✓
- 5. break a record ✓
crash a record
- 6. a railway stop
a railway station ✓
- 7. a car park ✓
a car parking
- 8. in the radio
on the radio ✓
- 9. cross the road ✓
cross through the road
- 10. get to home
get home ✓

11. do housework ✓
make housework
12. drive a horse
ride a horse ✓
13. tell a secret ✓
talk a secret
14. stop raining ✓
stop to rain
15. learn to drive ✓
learn driving
16. jump over the wall ✓
jump through the wall
17. at night ✓
in night
18. in the countryside ✓
on the countryside
19. sociable skills
social skills ✓
20. a small salary ✓
a little salary
21. make breakfast ✓
do breakfast
22. a boring job ✓
a bored job
23. pay for the bill
pay the bill ✓
24. the good bus
the right bus ✓
25. at lunchtime ✓
in lunchtime
26. a secure driver
a safe driver ✓
27. a driving test ✓
a drive test
28. a weak exercise
a gentle exercise ✓

29. a window cleaner ✓
a window brusher

30. late on class
late for class ✓

Task 2

Collocation accuracy judgments

Instructions:

Decide if the underlined phrases are correct or incorrect? If you think the phrase is incorrect, please give the correct phrase.

Here is an example.

1. I do exercise every day.
2. I live in the first floor.

You answer in the following way.

1. correct
2. incorrect; on the first floor

-
1. Do you know what is in TV tonight? (incorrect; on TV)
 2. Please answer the questions below the text. (correct)
 3. I hardly ever read in bed. (correct)
 4. He usually plays tennis in weekends. (incorrect; at weekends)
 5. What are you going to wear on the party? (incorrect; at the party)
 6. Do they work in a factory? (correct)
 7. He went through control passport an hour ago. (incorrect; passport control)
 8. His parents are waiting for him in the living room. (correct)
 9. Lorraine normally travels a car. (incorrect; travel by car)
 10. She writes a letter her sister every day. (incorrect; writes a letter to)
 11. They had a fast meal and left. (incorrect; a quick meal)
 12. The 5 bus takes you to the city centre. (correct)
 13. I was in Carlow in last week. (incorrect; last week)
 14. My parents make shopping twice a week. (incorrect; do/go shopping)
 15. Donald plays the guitar every now and then. (correct)

16. I need to get a few more Christmas postcards. (incorrect; Christmas cards)
17. I need to go to the cash machine to get some money. (correct)
18. She wants to win a medal next year. (correct)
19. I'm very tired after work. I have a 12-hour work day. (incorrect; working day)
20. I always get a bonus at Easter. (correct)
21. Will you make a photo of us, please? (incorrect; take a photo)
22. I hardly ever go on a walk in the forest. (incorrect; go for a walk)
23. My kids listen the radio every day. (incorrect; listen to the radio)
24. You can look that up in the Internet. (incorrect; on the Internet)
25. We'll go down to the sea on a swim before dinner. (incorrect; for a swim)
26. You should find the test pretty easy. (correct)
27. I'll be down in a minute. I'm just doing my hair. (correct)
28. Liz is currently working part-time only. (correct)
29. I've bought shampoo for regular washing. (correct)
30. What wonderful news! Congratulations! I am very happy for you. (correct)

Task 3

Fill-in-the-blank sentences

Instructions:

In each of the sentences there is ONE word missing. Please fill in each blank with a word that best completes each sentence.

Please do not guess if you do not know the answer.

Here is an example.

I live _____ the first floor.

You answer in the following way:

on

1. I'm having a job _____ on Friday. I hope it goes well, because I really want the job.
(interview)
2. Yesterday was a public _____ and all the banks were closed. (holiday)
3. No, he wasn't here. He was _____ work this time yesterday. (at)
4. Pupils usually _____ a uniform at school. (wear)
5. He was angry _____ everyone. Nobody really liked him. (at)
6. There's someone _____ the door. (at)

7. To get this job you need to know how to _____ a computer. (use)
8. Don't forget to _____ me a postcard when you're in Rome. (send)
9. I wanted to _____ on holiday with him but I changed my mind. (go)
10. Linda really enjoys _____ books. She's just finished "Alice in Wonderland". (reading)
11. I don't think I could work _____ an office. (in)
12. We had only one _____ of wine each, not the whole bottle. (glass)
13. I hardly ever _____ a bath. I prefer taking a shower. (have/take)
14. _____ the evening each of the men smoked a cigar. (in)
15. They arrived _____ the airport 2 hours before the scheduled flight. (at)
16. Let's _____ a drink to celebrate your success. (have)
17. It takes about 30 minutes _____ foot, or 10 minutes by bus. (on)
18. Yes, I remember that old lady who would always look out _____ the window. (of)
19. "Do you have your own house?" "No, I live in _____ accommodation." (rented)
20. You'd better _____ a doctor about that cough. (see/consult)
21. How many _____ of tea do you have a day? (cups)
22. The information is stored _____ the computer. (on)
23. He _____ a drawing of what our house looked like a few years ago. (made)
24. Did your parents give you pocket _____ when you were little. (money)
25. A lawyer is someone who represents his clients in _____. (court)
26. Niamh is moving up the _____ ladder quickly. She has been promoted yet again. (career)
27. Shall we go out _____ a meal and then on to a movie? (for)
28. I work from 9 to 5 _____ weekdays. (on)
29. She doesn't like wearing glasses. She's thinking of buying contact _____. (lenses)
30. Do you get a regular _____ rise at work? (pay)

B1 progress collocation test

Task 1
Multiple-choice questions

Instructions:

This is a collocation test. Please tick (✓) the correct phrase. Here is an example.

- C. do exercise
- D. make exercise

You answer it in the following way.

- C. do exercise ✓
 - D. make exercise
-

- 1. enter a competition ✓
enter a contract
- 2. tell a lie ✓
talk a lie
- 3. make a lot of activities
do a lot of activities ✓
- 4. accuse her of stealing ✓
accuse her about stealing
- 5. fierce competition ✓
devoted competition
- 6. a taxi stop
a taxi rank ✓
- 7. make big money ✓
make large money
- 8. congratulate them passing the exam
congratulate them on passing the exam ✓
- 9. busy roads ✓
trafficked roads
- 10. go on a travel
go on a trip ✓

11. play bowling
play bowls ✓
12. approve people to smoke
approve of people smoking ✓
13. in short supply ✓
in little supply
14. consist of five exercises ✓
consist with five exercises
15. avoid talking ✓
avoid to talk
16. be close a bus stop
be close to a bus stop ✓
17. let him go out late ✓
let him to go out late
18. a windy hall
a draughty hall ✓
19. make a slim fortune
make a small fortune ✓
20. plenty of space on storage
plenty of storage space ✓
21. dead straight ✓
dead bright
22. pour the bath
run the bath ✓
23. raise money ✓
rise money
24. be seriously rich ✓
be seriously employed
25. artificial move
false move ✓
26. a pedestrian sphere
a pedestrian precinct ✓
27. good to figures
good with figures ✓
28. lend him a book ✓
borrow him a book

29. a cramped house ✓
a tight house

30. throw a party ✓
make a party

Task 2

Collocation accuracy judgments

Instructions:

Decide if the underlined phrases are correct or incorrect? If you think the phrase is incorrect, please give the correct phrase.

Here is an example.

- 3. I do exercise every day.
- 4. I live in the first floor.

You answer in the following way.

- 3. correct
- 4. incorrect; on the first floor

-
- 1. This technology is bang up-to-date. (correct)
 - 2. I'd like to travel the world before I settle down. (correct)
 - 3. When I play football, my aim is to score a goal. (correct)
 - 4. I only saw the man for a split minute as he ran past. (incorrect; a split second)
 - 5. He tried to pick a fight with me. (correct)
 - 6. She attended in a lecture on migration. (incorrect; attended a lecture)
 - 7. Don't you get bored living out there on the suburbs? (incorrect; in the suburbs)
 - 8. When I have some sparing time, I will definitely have a look at your poems. (incorrect; spare time/time to spare)
 - 9. I can't make any phone calls. I'm out of credit. (correct)
 - 10. Despite all her hard work she dropped from university. (incorrect; dropped out of university)
 - 11. Mary has a very good memory. She can do complex mind arithmetic. (incorrect; mental arithmetic)
 - 12. My sister, Maggie, fell from the stairs when she was five and broke her arm. (incorrect; fell down the stairs)

13. He received his result exam yesterday. He failed the whole module. (incorrect; exam result)
14. He controlled himself by strict force of will. (incorrect; sheer force of will)
15. I don't know my way in the city yet. (incorrect; way around the city)
16. I am saving up to buy a small house. I live in a block at the moment. (incorrect; in a block of flats/ in an apartment block)
17. Go down this road and turn on the second traffic lights. (incorrect; at the second traffic lights)
18. Strong currents can be very dangerous for swimmers. (correct)
19. Short skirts are in fashion again. (correct)
20. The headlines were splashed across the front page of every newspaper. (correct)
21. I'm not really in speaking terms with him. (incorrect; on speaking terms)
22. Could you leave the keys to my office with the secretary? (correct)
23. Does the house have a view on the garden? (incorrect; a view of the garden)
24. Do you fancy a quick chat on the phone? (correct)
25. She looked rather lost and lonely, standing in a corner by herself. (correct)
26. You're too old to live in home with your parents. (incorrect; at home)
27. Hollywood discovered her and turned her in a star. (incorrect; turned her into a star)
28. Who should we invite to the party? (correct)
29. I am main provider and the chief of the household. (incorrect; the head of the household)
30. She graduated university and found a very good job. (incorrect; graduated from)

Task 3 Fill-in-the-blank sentences

Instructions:

In each of the sentences there is ONE word missing. Please fill in each blank with a word that best completes each sentence.

Please do not guess if you do not know the answer.

Here is an example.

I live _____ the first floor.

You answer in the following way:

on

1. He told the police a false _____ about being attacked, but they knew he was lying. (story)
2. During the ceremony the bride and the groom _____ the register. (signed)
3. We have just purchased a modern semi-_____ house. (detached)
4. Doctors have to _____ long hours. (work)
5. They invited us to _____ on holiday with them but it'd cost too much money. (go)
6. It was an unanimous _____ to scrap the project. (decision)
7. She is quite tall _____ her age. (for)
8. Jazz just isn't my cup of _____. I prefer classical music. (tea)
9. My house is situated in a quiet residential _____. (area/district)
10. If you want to make a profit, you need to think what to _____ your money in. Shares? The housing market? (invest)
11. She promised she would stay _____ contact but she didn't. (in)
12. Who does this jacket belong _____? (to)
13. The other _____ of his family were against the marriage. (members)
14. They _____ an argument last night. (had)
15. I don't think we have the same _____ of humour. (sense)
16. You've really got no need to worry _____ your weight. You're very slim. (about)
17. Brian gets on really _____ with his boss. (well)
18. At this stage the proposals are still subject _____ change. (to)
19. Don't let eating between meals _____ a habit. (become)
20. It normally takes me 20 minutes to get to work, but on that day I got to work in _____ time – 10 minutes tops. (record)
21. I would never lie to you. You can tell _____ my face that I'm telling the truth. (from)
22. Dan's crazy _____ extreme sports. (about)
23. All she is interested _____ is money. (in)
24. More people _____ suicide at Christmas than at any other time. Most of them die. (commit)
25. What are you going to spend money _____? (on)
26. I always faint _____ the sight of blood. (at)
27. Astronomers _____ a significant discovery about our galaxy last month. (made)
28. I'm looking for a _____ furnished flat – with fitted kitchen and furniture in other rooms. (fully)
29. I admire those who are good _____ writing. (at)
30. This is no _____ meeting. This meeting will decide about your future. (ordinary)

Table A.7. Words tested in the A2 and B1 Progress Vocabulary Tests and their mode of testing

Mode of testing	Test	Word	Part of speech	Receptive test	Productive test
receptive and productive	A2	alone	adj	23	16
		begin	verb	15	10
		boss	noun	28	1
		chef	noun	9	2
		collect	verb	18	11
		enjoy	verb	13	12
		expect	verb	17	13
		expensive	adj	2	17
		forget	verb	14	14
		jogging	noun	4	3
		meal	noun	12	4
		office	noun	26	5
		refuse	verb	16	15
		repetitive	adj	3	18
		sailing	noun	5	6
		shower	noun	27	7
		team	noun	29	8
		travel	noun	30	9
	B1	affair	noun	27	1
		approve	verb	17	10
		archery	noun	5	2
		attempt	verb	15	11
		blame	verb	18	12
		colleague	noun	9	3
		dependable	adj	21	16
		editor	noun	29	4
		embarrassed	adj	22	17
		fancy	verb	16	13
		proprietor	noun	28	5
		rise	verb	13	14
		striking	adj	19	18
		supplement	noun	12	6
		tabloid	noun	11	7
		trickster	noun	8	8
		urge	verb	14	15
		weakness	noun	25	9
receptive only	A2	accountant	noun	8	
		boring	adj	20	
		drawing	noun	6	
		friendly	adj	19	
		glamorous	adj	21	
		outdoor	adj	24	
		railway	noun	10	
		record	noun	11	
		salary	noun	25	
		sick	adj	22	
		undertaker	noun	7	
		wonderful	adj	1	
	B1	affordable	adj	24	
		article	noun	26	
		charm	noun	7	
		columnist	noun	30	
		confident	adj	2	
		flexible	adj	1	
		generous	adj	3	
		jogging	noun	6	
		opposite	adj	20	
		up-to-date	adj	23	
		wealth	noun	10	
		wrestling	noun	4	

Table A.8. The BBI types of collocations tested in the A2 Progress Collocation Test

Group of collocations	Type of phrase	Base form of the phrase	Task 1	Task 2	Task 3
Grammatical collocations	prep + noun (G4)	at lunchtime	25		
		at night	17		
		at the airport			15
		at the door			6
		at the party		5	
		at weekends		4	
		at work			3
		for a meal			27
		for a swim		25	
		for a walk		22	
		in a factory		6	
		in an office			11
		in bed		3	
		in court			25
		in the countryside	18		
		in the evening			14
		on a computer			22
		on foot			17
		on the Internet		24	
		on the radio	8		
		on TV		1	
		on weekdays			28
		out of the window			18
	adj + prep (G5)	angry at			5
		happy for sb/somebody		30	
		late for class	30		
	verb patterns (G8)				
		verb - allowing dative movement (G8-A)		10	
		verb + to-inf (G8-E)	15		9
		verb + prep (G8-D)		23	
		verb + prep (G8-D)		9	
		verb + prep (G8-D)		8	
		verb + gerund (G8-J)	14		
Lexical collocations	verb + noun (L1)	answer the questions		2	
		break a record	5		
		cross the road	9		
		do housework	11		
		do one's hair		27	
		do/go shopping		14	
		get a bonus		20	
		get home	10		
		go to (the party)	4		
		have a drink			16
		have/take a bath			13
		jump over the wall	16		
		leave the building	3		
		make a drawing			23
		make breakfast	21		
		pay the bill	23		
		play chess	2		
		play the guitar		15	
		read books			10
		ride a horse	12		
		see/consult a doctor			20
		send (sb) a postcard			8
		take a photo		21	
		tell a secret	13		

Group of collocations	Type of phrase	Base form of the phrase	Task 1	Task 2	Task 3
		use a computer wear a uniform win a medal		18	7 4
	adj + noun (L3)	boring job driving test fast meal gentle exercise last week public holiday regular washing rented accommodation right bus safe driver small salary social skills working day	22 27 28 24 26 20 19	11 13 29 19	2 19
	noun + (of) + noun (L5)	bar of soap car park career ladder cash machine Christmas cards city centre contact lenses cup of tea glass of wine job interview passport control pay rise pocket money railway stop window cleaner	1 7 6 29	17 16 12 7	26 29 21 12 1 30 24
	adverb of degree + adj (L6)	pretty easy		26	
	verb + adverb (L7)	work part-time		28	

Table A.9. The BBI types of collocations tested in the B1 Progress Collocation Test

Group of collocations	Type of phrase	Base form of the phrase	Task 1	Task 2	Task 3
Grammatical collocations	noun + prep (G1)	keys to (my office) view on (the garden)		22 23	
	prep + noun (G4)	at home at the sight at the traffic lights for sb's age in a corner in contact in fashion in the city in the suburbs on (speaking) terms on a trip on holiday out of credit	10	26 17 25 19 15 7 21 9	26 7 11 5
	adj + prep (G5)	close to crazy about	16		22

Group of collocations	Type of phrase	Base form of the phrase	Task 1	Task 2	Task 3
Lexical collocations	adj + noun (L3)	unanimous decision			6
	noun + (of) + noun (L5)	block of flats cup of tea exam result head of the household members of his family record time sense of humour storage space taxi rank	20 6	16 13 29	8 13 20 15
		adverb of degree + adj (L6)	21 24	1	28 17

APPENDIX H: Polish Vocabulary Test

Instructions:

W każdym zdaniu brakuje JEDNEGO słowa. Proszę uzupełnić brakujące słowa.

(In each of the sentences there is ONE word missing. Please fill in each blank with a word that best completes this sentence.)

1. Dokładnie, ale szybko. Jeśli będziecie zbyt opie_____ lub niedokładni, nie dostaniecie posiłku. Pracujecie tylko do zachodu słońca. (opieszali¹⁴⁰)
2. Pedro wypadł najlepiej w debacie telewizyjnej kandydatów. Wiedząc, że Meksykanie lubią typ zwycięzcy, Pedro afi_____ się ze swoją wielką pewnością siebie. (afiszował)
3. Duże ośrodki są przeładowane ogromną ilością spraw. Rodzi to powszechne niezadowolenie oraz korupcję, gdyż obr_____ przedsiębiorcy potrafią załatwić sobie wpis do rejestru w ciągu kilku czy kilkunastu dni, pozostali zaś czekają miesiącami. (obrotni)
4. Amerykańscy bot_____ uzyskali pierwsze odmiany uprawne w wyniku krzyżowania dzikich leśnych gatunków jagód. (botanicy)
5. Jedynym ustaleniem IX Soboru Watykańskiego było przyjęcie dog_____ o nieomyślności papieża. (dogmatu)
6. Dawno przestała im przygrywać ckl_____ melodyjka zakochanych. Romantyzm dawno się w ich związku skończył. (ckliwa)
7. Odnotowano również pierwsze próby plą_____ opuszczonych sklepów. Dwóch mężczyzn przyłapano na kradzieży zegarków. (plądrowania)
8. Marynarz odpowiedzialny za wskazania sonaru twierdził stanowczo, że jego uwagę tego dnia rozp_____ obecność kilkunastu cywilów. (rozproszyła)
9. Nastąpiło zmniejszenie liczby towarzystw, które są w trudnej sytuacji finansowej, z dziesięciu do sześciu. Nadzór nie chce ujawniać nazw spółek znajdujących się w tar_____. (tarapatach)
10. Pasażerowie autobusu wyszli na drogę, rozwinęli transparenty i całkowicie wst_____ ruch na obu pasach. (wstrzymali)

¹⁴⁰ English translations of the Polish target words are provided in alphabetical order in Table A.10 below the test.

11. W alk_____ było straszliwie duszno, toteż wieczorami Lucjan znosił prawdziwe męczarnie. Pościel parzyła mu ciało, a przed oczami kłębiły się kolorowe majaki. (alkowie)
12. Systematyczne odślanianie ruin dużego pałacu, z pięknymi moz_____ podłogowymi, ma miejsce w Nea Paphos na Cyprze. (mozaikami)
13. Każda fabryka, która zan_____ wodę pitną, podlega karze grzywny. (zanieczyszcza)
14. Zarządzona sekcja zwłok Karola nie wykazuje śladów ars_____ w jego organizmie. Maria zostaje więc aresztowana tylko pod zarzutem kradzieży, a nie morderstwa. (arszeniku)
15. Anna satyrą pa_____ się od 40 lat. Zajmuje się również malarstwem. (para)
16. Osoby, mieszkające w dużych aglomeracjach, są czasami pogardliwie nazywane mie_____. (mieszczuchami)
17. W rezultacie wprowadzonych zmian powstał nieopisany gali_____. Myliły się konkursy i filmy, a zasady przyporządkowania do jednej z dwóch kategorii były niezrozumiałe. (galimatias)
18. Studnia wiedzy jest dla wszystkich dostępna – tylko zan_____ ręce i czerpać. (zanurzyć)
19. W skw_____ popołudnie wypoczywaliśmy w sadzie, w cieniu pod lipą. (skwarne)
20. Sądzę, że stan wojenny pozostaje nadal w świadomości Polaków i tkwi jak bolesna, nie usunięta za_____ w narodowej pamięci. (zadra)
21. W zeszły poniedziałek minister opon_____ przeciwko temu werdyktowi jako niezgodnemu z prawem. (oponował)
22. Jan dotarł do jakiegoś drzewa i z całych sił chro_____ objął pień. Kurz powoli opadał, pojawiło się błękitne niebo. (chropowaty)
23. Każdy po stracie osoby najbliższej jest w głębokiej ża_____. Jednak coraz częściej pójście na groby zastępowane jest przez płatne ogłoszenie w gazecie. (żałobie)
24. Parę tygodni po pożarze żył jak w let_____, przygnębiony, nie mogąc się odnaleźć. Któregoś dnia obudził się z odrętwienia. Postanowił, że odbuduje swój dom. (letargu)
25. Niebieski ży_____ jasno oświeślał pokój i uwydatniał rysy na meblach. (żyrandol)
26. Gdy Nowak w lokalu kontaktowym składał relację, funkcjonariusz SB ba_____ coś w pośpiechu na kartce. (bazgrał)
27. Rodzice jednej z dziewczynek zauważyli, że dziecko jest oso_____, senne i nienaturalnie spokojne. Po zbadaniu dziewczynki okazało się, że podano jej silny środek uspokajający. (osowiale)
28. Pubów powstaje coraz więcej. Wzd_____ się na samo słowo "pub". Puby zostawmy Anglikom. My mamy piwiarnie, karczmy, gospody – są to dobre, tradycyjne nazwy. (wzdrygam)

29. Zebrani na szczycie w Berlinie, prowadzili wczoraj nie kończące się dyskusje w kiluosobowych grupach, aby przełamać im _____ w debacie nad wewnętrznymi reformami w UE. Doszli w końcu do porozumienia. (impas)
30. Dla aktorki niezmiennie dziewczęcy głos albo zbyt młoda apa _____ może być problemem, jak to miało miejsce w przypadku Anny Dymnej. Jako dwudziestokilkuletnia kobieta wciąż wyglądała jak nastolatka. (aparycja)
31. Nieoficjalnie dowiedzieliśmy się, że Duńczycy fin _____ obecnie transakcję z Amerykanami. Strony oczekują jeszcze tylko na zgodę od ministra finansów. (finalizują)
32. Ma _____ ośmiornicy opłoty ofiarę na ekranie telewizora. (macki)
33. Był wiatr, więc silnik było słabiej słychać. W dzień bezwietrzny wa _____ silnika byłby z pewnością jeszcze bardziej słyszalny. (warkot)
34. Uznała ich za niebezpiecznych barb _____, bo ich prawo nakazywało, by odrąbywać ręce złodziejom. (barbarzyńców)
35. Przestań się wie _____! Nie możesz przez 5 minut spokojnie usiedzieć na miejscu? (wiercić)
36. Na czele polskiej delegacji stanie premier Jerzy Buzek. Uroczystości tra _____ będzie telewizyjna Jedyńka. (transmitować)
37. Ten atak nie pozostawia żadnych wątpliwości co do zamiarów Waszyngtonu, który obecnie uz _____ sobie rolę światowego żandarma. (uzurpuje)
38. Jennifer Lopez wcieli się w cudz _____ i alkoholiczkę w kinowej adaptacji popularnego serialu "Dallas". Swojego ekranowego męża zdradzać będzie niejednokrotnie. (cudzołożnicę)
39. "Polak" nie ma w naszym języku pej _____ znaczenia, ale w języku polskim słowo "Żyd" bardzo często jest obelgą. (pejoratywnego)
40. Miałem w sądzie sprawę ze spółdzielnią mieszkaniową. Sąd wydał wyrok, w którym zasądził pewną kwotę na moją rzecz. Wyrok ten musi się upr _____. (uprawomocnić)
41. W sklepie tym zakupić można her _____ pojemniki na produkty spożywcze. Pojemniki te nie przepuszczają powietrza. (hermetyczne)
42. Za godzinę będzie zupełnie ciemno. Fir _____ nieba roziskrzy się tysiącami gwiazd. (firmament)
43. Aby uwi _____ kopie wpłaty należy udać się do notariusza. (uwierzytelnić)
44. Mariola to wier _____ kłamczucha i jest pewna, że jej piękna buźka i łzy zmylą każdego. (wierutna)
45. Co mnie zaskakuje w Tajlandii to ogromna pru _____. Kwitnie tam przemysł erotyczny, ale nie wolno wwozić do kraju żadnej pornografii, a z prezentowanych w telewizji filmów amerykańskich wycina się wszelkie mocniejsze sceny. (pruderyjność)

46. Komentarza wymaga już sam fakt, że dla wykonania czynności tak nieskomplikowanej jak pomiar trzeba kons_____ specjalne teorie. (konstruować)
47. Kierowano ich do pracy od świtu do nocy jako par_____, niewolników, a miejscem ich spania było to, co dzisiaj możemy nazwać oborą, chlewnią czy jakimkolwiek poddaszem. (parobków)
48. Kiedy powiedziała mu, że jest w ciąży, zwinął ma_____ i uciekł. (manatki)
49. Na lewo od kontuaru znajdowała się winda. Wybita była ciemnoczerwonym suk_____, już dobrze spłóviałym, miejscami zaś przetartym. (suknem)
50. W zeszłym tygodniu król Kambodży rat_____ ustawę o powołaniu trybunału do osądzenia Czerwonych Khmerów. (ratyfikował)
51. Pobrali się dopiero późną zimą. Rodzice początkowo odmawiali błogosławieństwa ze względu na mez_____. On pochodził z bardzo biednej rodziny. Ona była arystokratką. (mezalians)
52. Istnieje też cała gama dostępnych specjalistom technik badawczych, pozwalających na demi_____ fałszerstwa, np. zdjęcia rentgenowskie, czy badania mikroskopowe. (demistyfikację)
53. Wysoka Izbo, w przypadku tej ustawy należy wziąć pod uwagę czas roz_____ dziecka z rodziną i fakt pozostawiania tego dziecka przez bardzo długi czas w domu dziecka czy w placówce wychowawczo-opiekuńczej. (rozłąki)
54. Przestań świecić mi latarką w oczy. Od tego można dostać oco_____! (oczopląsu)

Table A.10. English translations of Polish words from the Polish Vocabulary Test

Base forms	Part of speech	Translation	No on the Polish Vocabulary Test
afiszować	verb	flaunt	2
alkowa	noun	alcove, small chamber	11
aparycja	noun	appearance	30
arszenik	noun	arsenic	14
barbarzyńca	noun	barbarian	34
bazgrać	verb	scribble	26
botanic	noun	botanist	4
chropowaty	adj	coarse	22
ckliwy	adj	mawkish, maudlin	6
cudzołożnica	noun	adulteress	38
demistyfikacja	noun	demystification	52
dogmat	noun	dogma	5
finalizować	verb	finalize	31
firmament	noun	firmament	42
galimatias	noun	hotchpotch	17
hermetyczny	adj	airtight, hermetic	41
impas	noun	deadlock, impasse	29
konstruować	verb	construct	46
letarg	noun	lethargy	24
macka	noun	tentacle	32
manatki	noun	clobber	48
mezalians	noun	marriage beneath one	51

Base forms	Part of speech	Translation	No on the Polish Vocabulary Test
mieszczuch	noun	city-slicker	16
mozaika	noun	mosaic	12
obrotny	adj	resourceful, savvy	3
oczopląs	noun	the feeling of being confused or overwhelmed as a result of vibrant colours, etc.	54
opieszwały	adj	sluggish, inert	1
oponować	verb	oppose	21
osowiały	adj	listless	27
parobek	noun	farmhand	47
parać (się)	verb	earn one's living doing something, dabble in something	15
pejoratywny	adj	pejorative	39
plądrowanie	noun	plunder, looting	7
pruderyjność	noun	prudishness	45
ratyfikować	verb	ratify	50
rozłąka	noun	separation (e.g. of families)	53
rozproszyc	verb	scatter, disperse	8
skwarny	adj	scorching hot	19
sukno	noun	cloth	49
tarapaty	noun	plight, trouble	9
transmitować	verb	televisé, broadcast	36
uprawomocnić	verb	validate (of a ruling)	40
uwierzytelnić	verb	authenticate	43
uzurpować	verb	usurp	37
wiercić (się)	verb	fidget, squirm	35
wierutny	adj	downright (of a lie)	44
wstrzymać	verb	halt (of traffic)	10
wzdrygać (się)	verb	shiver, cringe	28
zadra	noun	grudge	20
zanieczyszczać	verb	contaminate	13
zanurzyć	verb	dip (e.g. hands in the water)	18
żałoba	noun	mourning, bereavement	23
warkot	noun	whirr	33
żyrandol	noun	chandelier	25

Table A.11. Frequencies of occurrence of target Polish words in the IPI PAN corpus

Part of speech	Base forms	Overall frequency of inflected forms	No on the Polish Vocabulary Test
adjectives	osowiały ¹⁴¹	35	27
	skwarny	36	19
	ckliwy	50	6
	wierutny	75	44
	chropowaty	96	22
	obrotny	146	3
	opieszaly	146	1
	pejoratywny	174	39
nouns	hermetyczny	316	41
	pruderyjność	1	45
	demistyfikacja	4	52
	oczopląs	15	54
	cudzołożnica	25	38
	alkowa	39	11
	mezalians	45	51
	plądrowanie	53	7
	manatki	60	48
	zadra	62	20
	aparycja	63	30
	firmament	79	42
	arszenik	80	14
	galimatias	111	17
	warkot	124	33
	impas	141	29
	macka	149	32
	żyrandol	164	25
	mieszczuch	172	16
	letarg	198	24
	rozłąka	203	53
	barbarzyńca	256	34
	sukno	304	49
	dogmat	375	5
	parobek	387	47
	tarapaty	433	9
	botanic	463	4
	mozaika	527	12
	żałoba	571	23
verbs	bazgrać	22	26
	uwierzytelnić	25	43
	wzdrygać (się)	37	28
	wiercić (się)	56	35
	afiszować	61	2
	uzurpować	219	37
	uprawomocnić	275	40
	zanieczyszczać	280	13
	parać (się)	339	15
	transmitować	359	36
	konstruować	380	46
	oponować	442	21
	ratyfikować	446	50
	zanurzyć	464	18
	rozproszyc	515	8
	finalizować	524	31
	wstrzymać	549	10

¹⁴¹ English translations of the Polish target words are presented in Table A.10.

Table A.12. Frequencies of occurrence of target Polish verbs and their inflected forms in the IPI PAN corpus – alphabetical order

Base forms (verbs)	Inflected forms	Frequencies of inflected forms	Overall frequency	No on the Polish Test
afiszować ¹⁴²	afiszować [inf:imperf] ¹⁴³	31	61	2
	afiszował [sg:m:imperf]	8		
	afiszowała [sg:f:imperf]	1		
	afiszowali [pl:m:imperf]	2		
	afiszowały [pl:f:imperf]	1		
	afiszuj [impt:sg:2nd:imperf]	2		
	afiszują [pl:3rd:imperf]	7		
	afiszuje [sg:3rd:imperf]	8		
bazgrać	afiszujemy [pl:1st:imperf]	1	22	26
	bazgrać [inf:imperf]	2		
	bazgrał [sg:m:imperf]	1		
	bazgrali [pl:m:imperf]	1		
	bazgrzą [pl:3rd:imperf]	2		
	bazgrzę [sg:1st:imperf]	1		
	bazgrze [sg:3rd:imperf]	4		
	nabazgrać [inf:perf]	2		
finalizować	nabazgrał [sg:m:perf]	6	524	31
	nabazgrała [sg:f:perf]	3		
	finalizować [inf:imperf]	53		
	finalizował [sg:m:imperf]	35		
	finalizowała [sg:f:imperf]	7		
	finalizowali [pl:m:imperf]	14		
	finalizowało [sg:n:imperf]	2		
	finalizowały [pl:f:imperf]	4		
	finalizuj [impt:sg:2nd:imperf]	3		
	finalizują [pl:3rd:imperf]	37		
	finalizując [advpart:imperf]	199		
	finalizuję [sg:1st:imperf]	4		
	finalizuje [sg:3rd:imperf]	90		
konstruować	finalizujemy [pl:1st:imperf]	68		
	finalizujesz [sg:2nd:imperf]	8		
	konstruować [inf:imperf]	77	380	46
	konstruował [sg:m:imperf]	34		
	konstruowała [sg:f:imperf]	3		
	konstruowali [pl:m:imperf]	47		
	konstruowało [sg:n:imperf]	2		
	konstruowały [pl:f:imperf]	5		
	konstruują [pl:3rd:imperf]	18		
	konstruując [advpart:imperf]	48		
	konstruuję [sg:1st:imperf]	6		
	konstruuje [sg:3rd:imperf]	56		
	konstruujecie [pl:2nd:imperf]	1		
	konstruujemy [pl:1st:imperf]	4		
	skonstruować [inf:perf]	42		
	skonstruował [sg:m:perf]	20		
	skonstruowała [sg:f:perf]	4		
	skonstruowali [pl:m:perf]	8		
	skonstruuję [sg:1st:perf]	1		
	skonstruuje [sg:3rd:perf]	1		
	skonstruujemy [pl:1st:perf]	2		
oponować	skonstruujmy [impt:pl:1st:perf]	1	442	21
	oponować [inf:imperf]	51		
	oponował [sg:m:imperf]	124		

¹⁴² English translations of the Polish target words are presented in Table A.10.

¹⁴³ perf, imperf – perfective, imperfective (aspect)
sg, pl – singular, plural (number)
1st, 2nd, 3rd – first, second, third (person)
impt – imperative (mode)
m, f, n – masculine, feminine and neuter (gender)
advpart – adverbial participle

Base forms (verbs)	Inflected forms	Frequencies of inflected forms	Overall frequency	No on the Polish Test
	oponowała [sg:f:imperf] oponowali (pl:m:imperf) oponowało [sg:n:imperf] oponowały [pl:f:imperf] [pl:n:imperf] oponują [pl:3rd:imperf] oponując [advpart:imperf] oponuję [sg:1st:imperf] oponuje [sg:3rd:imperf] oponujemy [pl:1st:imperf] zaoponować [inf:perf] zaoponował [sg:m:perf] zaoponowała [sg:f:perf] zaoponowali [pl:m:perf] zaoponuje [sg:3rd:perf]	26 35 3 4 4 37 5 14 72 8 10 29 16 7 1		
parać (się)	para [sg:3rd:imperf] parać [inf:imperf] paracie [pl:2nd:imperf] parają [pl:3rd:imperf] parając [advpart:imperf] parał [sg:m:imperf] parała [sg:f:imperf] parali [pl:m:imperf] parało [sg:n:imperf] parały [pl:f:imperf][pl:n:imperf] param [sg:1st:imperf] paramy [pl:1st:imperf] paraszą [sg:2nd:imperf]	91 32 1 49 3 84 8 40 5 7 14 3 2	339	15
ratyfikować	ratyfikować [inf:perf][inf:imperf] ratyfikował [sg:m:imperf] [sg:m:perf] ratyfikowała [sg:f:imperf][sg:f:perf] ratyfikowali [pl:m:perf] [pl:m:imperf] ratyfikowało [sg:n:imperf] [sg:n:perf] ratyfikowały [pl:f:perf][pl:f:imperf] ratyfikują [pl:ter:imperf][pl:ter:perf] ratyfikując [advpart:perf] [advpart:imperf] ratyfikuję [sg:1st:perf] [sg:1st:imperf] ratyfikuje [sg:3rd:imperf] [sg:3rd:perf] ratyfikujemy [pl:1st:imperf] [pl:1st:perf]	77 62 91 14 18 35 11 70 1 57 10	446	50
rozproszyć	rozprasza [sg:1st:imperf] rozpraszać [inf:imperf] rozpraszacie [pl:2nd:imperf] rozpraszaj [sg:2nd:imperf] rozpraszają [pl:3rd:imperf] rozpraszając [advpart:imperf] rozpraszał [sg:m:imperf] rozpraszała [sg:f:imperf] rozpraszali [pl:m:imperf] rozpraszało [sg:n:imperf] rozpraszały [pl:f:imperf] rozpraszam [sg:1st:imperf] rozpraszamy [pl:1st:imperf] rozpraszasz [sg:2nd:imperf] rozprosz [impt:sg:2nd:perf]	109 71 2 4 34 14 16 13 9 20 17 3 8 2 2	515	8

Base forms (verbs)	Inflected forms	Frequencies of inflected forms	Overall frequency	No on the Polish Test
	rozproszą [pl:3rd:perf]	4		
	rozproszcie [impt:pl:2nd:perf]	1		
	rozproszę [sg:1st:perf]	1		
	rozproszy [sg:3rd:perf]	12		
	rozproszyć [inf:perf]	68		
	rozproszycie [pl:2nd:perf]	1		
	rozproszył [sg:m:perf]	28		
	rozproszyła [sg:f:perf]	22		
	rozproszyli [pl:m:perf]	15		
	rozproszyło [sg:n:perf]	14		
	rozproszyły [pl:f:perf]	22		
	rozproszymy [pl:1st:perf]	1		
	rozproszysz [sg:2nd:perf]	2		
transmitować	transmitować [inf:imperf]	111	359	36
	transmitował [sg:m:imperf]	16		
	transmitowała [sg:f:imperf]	72		
	transmitowali [pl:m:imperf]	7		
	transmitowało [sg:n:imperf]	15		
	transmitowały [pl:f:imperf]	16		
	transmituj [impt:sg:2nd:imperf]	2		
	transmitują [pl:3rd:imperf]	11		
	transmitując [advpart:imperf]	8		
	transmituję [sg:1st:imperf]	2		
	transmituje [sg:3rd:imperf]	96		
	transmitujemy [pl:1st:imperf]	3		
uprawomocnić	uprawomocnia [sg:3rd:imperf]	16	275	40
	uprawomocniać [inf:imperf]	1		
	uprawomocniają [pl:3rd:imperf]	8		
	uprawomocniając [advpart:imperf]	1		
	uprawomocniał [sg:m:imperf]	2		
	uprawomocniała [sg:f:imperf]	2		
	uprawomocniało [sg:n:imperf]	1		
	uprawomocniały [pl:m:imperf]	3		
	uprawomocni [sg:3rd:perf]	69		
	uprawomocnią [pl:3rd:perf]	7		
	uprawomocnić [inf:perf]	42		
	uprawomocnił [sg:m:perf]	38		
	uprawomocniła [sg:f:perf]	37		
	uprawomocniło [sg:n:perf]	30		
	uprawomocniły [pl:f:perf]	18		
uwierzytelnić	uwierzytelnia [sg:3rd:imperf]	4	25	43
	uwierzytelniać [inf:imperf]	1		
	uwierzytelniają [pl:3rd:imperf]	1		
	uwierzytelniając [advpart:imperf]	1		
	uwierzytelniał [sg:m:imperf]	1		
	uwierzytelniała [sg:f:imperf]	1		
	uwierzytelnić [inf:perf]	14		
	uwierzytelnił [sg:m:perf]	2		
uzurpować	uzurpować [inf:perf] [inf:imperf]	31	219	37
	uzurpował [sg:m:imperf]			
	[sg:m:perf]	24		
	uzurpowała [sg:f:imperf]			
	[sg:f:perf]	5		
	uzurpowali [pl:m:perf]			
	[pl:m:imperf]	13		
	uzurpowało [sg:n:imperf]			
	[sg:n:perf]	4		
	uzurpowały [pl:f:perf]			
	[pl:f:imperf]	5		
	uzurpują [pl:3rd:imperf]			
	[pl:3rd:perf]	22		
	uzurpując [advpart:perf]			
	[advpart:imperf]	15		

Base forms (verbs)	Inflected forms	Frequencies of inflected forms	Overall frequency	No on the Polish Test
	uzurpuję [sg:1st:perf] [sg:1st:imperf]	14		
	uzurpuje [sg:3rd:imperf] [sg:3rd:perf]	77		
	uzurpujecie [pl:2nd:imperf] [pl:2nd:perf]	1		
	uzurpujemy [pl:1st:imperf] [pl:1st:perf]	6		
	uzurpujmy [impt:pl:1st:imperf] [impt:pl:1st:perf]	2		
wiercić (się)	wiercą [pl:3rd:imperf] wiercąc [advpart:imperf] wierci [sg:3rd:imperf] wiercić [inf:imperf] wiercił [sg:m:imperf] wierciła [sg:f:imperf] wiercili [pl:m:imperf] wierciło [sg:n:imperf] wierciły [pl:f:imperf] wiercimy [pl:1st:imperf] wiercisz [sg:sec:imperf]	3 2 14 9 17 5 1 1 2 1 1	56	35
wstrzymać	wstrzymują [pl:3rd:imperf] wstrzymując [advpart:imperf] wstrzymuję [sg:1st:imperf] wstrzymuje [sg:3rd:imperf] wstrzymujecie [pl:sec:imperf] wstrzymujemy [pl:1st:imperf] wstrzymujesz [sg:2nd:imperf] wstrzymywać [inf:imperf] wstrzymywał [sg:m:imperf] wstrzymywała [sg:f:imperf] wstrzymywali [pl:m:imperf] wstrzymywało [sg:n:imperf] wstrzymywały [pl:f:imperf] wstrzymać [inf:perf]	42 43 11 66 1 29 1 53 37 23 24 28 8 183	549	10
wzdrygać (się)	wzdryga [sg:3rd:imperf] wzdrygać [inf:imperf] wzdrygają [pl:3rd:imperf] wzdrygając [advpart:imperf] wzdrygał [sg:m:imperf] wzdrygali [pl:m:imperf] wzdrygało [sg:n:imperf] wzdrygała [sg:f:imperf] wzdrygały [pl:f:imperf] wzdrygam [sg:1st:imperf] wzdrygamy [pl:1st:imperf]	9 2 3 2 9 1 1 4 3 2 1	37	28
zanieczyszczać	zanieczyszcza [sg:3rd:imperf] zanieczyszczać [inf:imperf] zanieczyszczają [pl:3rd:imperf] zanieczyszczając [advpart:imperf] zanieczyszczał [sg:m:imperf] zanieczyszczała [sg:f:imperf] zanieczyszczeni [pl:m:imperf] zanieczyszczały [pl:f:imperf] zanieczyszczasz [sg:2nd:imperf] zanieczyszcisci [sg:3rd:perf] zanieczyszcisci [inf:perf] zanieczyszciscił [sg:m:perf] zanieczyszcisciła [sg:f:perf] zanieczyszciscili [pl:m:perf] zanieczyszcisciły [pl:f:perf] zanieczyszciscimy [pl:3rd:perf] zanieczyszczą [pl:3rd:perf] zanieczyszczę [sg:1st:perf]	75 21 103 15 14 5 4 13 2 2 9 7 2 1 3 1 2 1	280	13

Base forms (verbs)	Inflected forms	Frequencies of inflected forms	Overall frequency	No on the Polish Test
zanurzyć	zanurza [sg:3rd:imperf]	53	464	18
	zanurzać [inf:imperf]	26		
	zanurzacie [pl:2nd:imperf]	1		
	zanurzaj [impt:sg:2nd:imperf]	1		
	zanurzają [pl:3rd:imperf]	14		
	zanurzając [advpart:imperf]	23		
	zanurzał [sg:m:imperf]	20		
	zanurzała [sg:f:imperf]	11		
	zanurzali [pl:m:imperf]	9		
	zanurzało [sg:n:imperf]	3		
	zanurzały [pl:f:imperf]	6		
	zanurzam [sg:1st:imperf]	10		
	zanurzamy [pl:1st:imperf]	15		
	zanurzasz [sg:2nd:imperf]	2		
	zanurz [impt:sg:2nd:perf]	4		
	zanurzą [pl:3rd:perf]	8		
	zanurzcie [impt:pl:2nd:perf]	1		
	zanurzę [sg:1st:perf]	1		
	zanurzmy [impt:pl:1st:perf]	4		
	zanurzy [sg:3rd:perf]	12		
	zanurzyć [inf:perf]	121		
	zanurzycie [pl:2nd:perf]	2		
	zanurzył [sg:m:perf]	68		
	zanurzyła [sg:f:perf]	18		
	zanurzyli [pl:m:perf]	16		
	zanurzyło [sg:n:perf]	6		
	zanurzyły [pl:f:perf]	3		
	zanurzymy [pl:1st:perf]	6		

Table A.13. Frequencies of occurrence of target Polish nouns and their inflected forms in the IPI PAN corpus – alphabetical order

Base forms (nouns)	Inflected forms	Frequency of inflected forms	Overall frequency	No on the Polish Test
alkowa ¹⁴⁴	alków [pl:gen:f] ¹⁴⁵	4	39	11
	alkową [sg:inst:f]	2		
	alkowa [sg:nom:f]	3		
	alkowę [sg:acc:f]	2		
	alkowie [sg:loc:f]	7		
	alkowy [pl:nom:f][sg:gen:f]	21		
aparycja	aparycją [sg:inst:f]	12	63	30
	aparycja [sg:nom:f]	17		
	aparycję [sg:acc:f]	12		
	aparycji [sg:gen:f][sg:loc:f]	22		
arszenik	arszenik [sg:acc:m][sg:nom:m]	34	80	14
	arszenikiem [sg:inst:m]	23		
	arszeniku [sg:gen:m]	23		
barbarzyńca	barbarzyńcą [sg:inst:m]	19	256	34
	barbarzyńcach [pl:loc:m]	3		
	barbarzyńcami [pl:inst:m]	29		
	barbarzyńcę [sg:acc:m]	7		
	barbarzyńcom [pl:dat:m]	7		
	barbarzyńców [pl:gen:m]	89		
	barbarzyńcy [sg:gen:m]	102		

¹⁴⁴ English translations of the Polish target words are presented in Table A.10..

¹⁴⁵ sg, pl – singular, plural (number)

m, f, n – masculine, feminine and neuter (gender)

nom, gen, dat, acc, inst, loc, voc – nominative, genitive, dative, accusative, instrumental, locative and vocative (cases)

Base forms (nouns)	Inflected forms	Frequency of inflected forms	Overall frequency	No on the Polish Test
botanik	botanicy [pl:nom:m] botanik [sg:nom:m] botanika [sg:acc:m][sg:gen:m] botanikami [pl:inst:m] botanikiem [sg:inst:m] botanikom [pl:dat:m] botaników [pl:acc:m][pl:gen:m] botanikowi [sg:dat:m]	113 106 18 5 10 4 206 1	463	4
cudzołożnica	cudzołożnic [pl:gen:f] cudzołożnica [sg:nom:f] cudzołożnicami [pl:inst:f] cudzołożnice [pl:acc:f] cudzołożnicę [sg:acc:f] cudzołożnicy [sg:gen:f]	14 1 1 1 5 3	25	38
demistyfikacja	demistyfikacja [sg:nom:f] demistyfikacji [sg:gen:f] demistyfiacją [sg:inst:f]	1 2 1	4	52
dogmat	dogmacie [sg:loc:m] dogmat [sg:acc:m][sg:nom:m] dogmatami [pl:inst:m] dogmatem [sg:inst:m] dogmatom [pl:dat:m] dogmatów [pl:gen:m] dogmatowi [sg:dat:m] dogmatu [sg:gen:m] dogmaty [pl:acc:m]	6 104 6 16 42 7 62 5 93 34	375	5
firmament	firmamencie [sg:loc:m] firmament [sg:acc:m][sg:nom:m] firmamentem [sg:inst:m] firmamentu [sg:gen:m]	53 9 2 15	79	42
galimatias	galimatias [sg:acc:m][sg:nom:m] galimatiasem [sg:inst:m] galimatiasie [sg:loc:m] galimatiasowi [sg:dat:m] galimatiasu [sg:gen:m]	81 3 8 1 18	111	17
impas	impas [sg:acc:m] impas [sg:nom:m]	66 75	141	29
letarg	letarg [sg:acc:m][sg:nom:m] letargu [sg:gen:m][sg:loc:m]	33 165	198	24
macka	macek [pl:gen:f] macką [sg:inst:f] macka [sg:nom:f] mackach [pl:loc:f] mackami [pl:inst:f] mackę [sg:acc:f] macki [pl:acc:f][pl:nom:f][sg:gen:f]	72 3 10 1 14 1 48	149	32
manatki	manatki [pl:acc:n][pl:nom:n] manatków [pl:gen:n]	58 2	60	48
mezalians	mezalians [sg:acc:m][sg:nom:m] mezaliansami [pl:inst:m] mezaliansie [pl:nom:m] mezaliansem [sg:inst:m] mezaliansów [pl:gen:m] mezaliansu [sg:gen:m] mezaliansy [pl:nom:m]	23 1 2 7 1 10 1	45	51
mieszczuch	mieszczuch [sg:nom:m] mieszczucha [sg:acc:m][sg:gen:m] mieszczuchach [pl:loc:m] mieszczuchami [pl:inst:m] mieszczuchem [sg:inst:m] mieszczuchom [pl:dat:m]	25 16 1 2 3 18	172	16

Base forms (nouns)	Inflected forms	Frequency of inflected forms	Overall frequency	No on the Polish Test
	mieszczuchów [pl:acc:m][pl:gen:m] mieszczuchowi [sg:dat:m] mieszczuchu [sg:loc:m] mieszczuchy [pl:nom:m]	64 2 2 39		
mozaika	mozaice [sg:loc:f] mozaik [pl:gen:f] mozaiką [sg:inst:f] mozaika [sg:nom:f] mozaikach [pl:loc:f] mozaikami [pl:inst:f] mozaikę [sg:acc:f] mozaiki [pl:acc:f][pl:nom:f][sg:gen:f] mozaikom [pl:dat:f]	28 37 45 121 10 17 96 171 2	527	12
oczopląs	oczopląs [sg:acc:m][sg:nom:m] oczopląsów [pl:gen:m] oczopląsu [sg:gen:m] oczopląsy [pl:nom:m]	7 1 6 1	15	54
parobek	parobcy [pl:nom:m] parobek [sg:nom:m] parobka [sg:acc:m][sg:gen:m] parobkach [pl:loc:m] parobkami [pl:inst:m] parobki [pl:nom:m] parobkiem [sg:inst:m] parobkom [pl:dat:m] parobków [pl:acc:m][pl:gen:m] parobkowi [sg:dat:m] parobku [sg:loc:m][sg:voc:m]	41 89 64 2 23 23 23 9 92 8 13	387	47
plądrowanie	plądrowania [sg:gen:n:imperf] plądrowanie [sg:acc:n:imperf][sg:nom:n:imperf] plądrowaniem [sg:inst:n:imperf] plądrowaniu [sg:dat:n:imperf]	29 14 3 7	53	7
pruderyjność	pruderyjność [sg:nom:f]	1	1	45
rozłąka	rozłące [sg:loc:f] rozłak [pl:gen:f] rozłaką [sg:inst:f] rozłaka [sg:nom:f] rozłakę [sg:acc:f] rozłaki [pl:acc:f][pl:nom:f][sg:gen:f]	24 2 10 40 51 76	203	53
sukno	sukna [pl:acc:n][pl:nom:n] [sg:gen:n] suknach [pl:loc:n] suknami [pl:inst:n] suknem [sg:inst:n] suknie [sg:loc:n] sukno [sg:acc:n][sg:nom:n]	130 2 2 71 18 81	304	49
tarapaty	tarapaty [pl:nom:n] tarapatów [pl:gen:n] tarapatami [pl:inst:n] tarapatach [pl:loc:n]	231 54 3 145	433	9
zadra	zadr [pl:gen:f] zadrą [sg:inst:f] zadra [sg:nom:f] zadrach [pl:loc:f] zadrami [pl:inst:f] zadrę [sg:acc:f] zadry [pl:acc:f][pl:nom:f][sg:gen:f]	2 7 38 1 1 4 9	62	20
żałoba	żałobą [sg:inst:f]	34	571	23

Base forms (nouns)	Inflected forms	Frequency of inflected forms	Overall frequency	No on the Polish Test
	żałoba [sg:nom:f] żałobami [pl:inst:f] żałobę [sg:acc:f] żałobie [sg:dat:f][sg:loc:f] żałoby [pl:acc:f][pl:nom:f][sg:gen:f]	110 1 82 94 250		
warkot	warkocie [sg:loc:m] warkot [sg:acc:m][sg:nom:m] warkotem [sg:inst:m] warkotów [pl:gen:m] warkotu [sg:gen:m] warkoty [pl:nom:m]	3 91 15 1 13 1	124	33
żyrandol	żyrandol [sg:acc:m][sg:nom:m] żyrandola [sg:gen:m] żyrandolach [pl:loc:m] żyrandolami [pl:inst:m] żyrandole [pl:acc:m][pl:nom:m] żyrandolem [sg:inst:m] żyrandoli [pl:gen:m] żyrandolów [pl:gen:m] żyrandolu [sg:loc:m]	50 15 3 5 52 12 24 1 2	164	25

Table A.14. Frequencies of occurrence of target Polish adjectives and their inflected forms in the IPI
PAN corpus – alphabetical order

Base forms (adjectives)	Inflected forms	Frequency of inflected forms	Overall frequency	No on the Polish Test
chropowaty ¹⁴⁶	chropowatą [sg:acc:f] ¹⁴⁷	12	96	22
	chropowata [csg:nom:f]	15		
	chropowate [pl:acc:f][pl:nom:f]	20		
	chropowatego [sg:gen:m][sg:gen:n]	3		
	chropowatej [sg:dat:f][sg:loc:f]	11		
	chropowaty [sg:acc:m][sg:nom:m]	15		
	chropowatych [pl:gen:f][pl:gen:m]			
	[pl:gen:n]	17		
ckliwy	chropowatym [sg:inst:m]	3	50	6
	ckliwą [sg:inst:f]	1		
	ckliwa [sg:nom:f]	5		
	ckliwe [pl:acc:f][pl:acc:m][pl:acc:n]	15		
	ckliwego [sg:gen:m]	4		
	ckliwej [sg:gen:f]	1		
	ckliwy [sg:acc:m][sg:nom:m]	10		
	ckliwych			
hermetyczny	[pl:gen:f][pl:gen:m][pl:gen:n]	7	316	41
	ckliwym [pl:dat:n]	7		
	hermetyczną [sg:acc:f]	34		
	hermetyczna [sg:nom:f]	37		
	hermetyczne [pl:acc:f]	75		
	hermetycznego [sg:gen:m]	20		
	hermetycznej [sg:gen:f]	17		
	hermetyczni [pl:nom:m]	4		
obrotny	hermetyczny [sg:acc:m][sg:nom:m]	50	146	3
	hermetycznych [pl:gen:f][pl:loc:f]	29		
	hermetycznym			
	[sg:inst:m][sg:inst:m]	50		
	obrotną [sg:acc:f][sg:inst:f]	2		
	obrotna [sg:nom:f]	7		
	obrotne [pl:nom:f][pl:nom:n]	7		
	obrotnego [sg:acc:m][sg:gen:m]	12		
opieszali	obrotnej [sg:gen:f]	1	146	1
	obrotnemu [sg:dat:m]	3		
	obrotni [pl:nom:m]	29		
	obrotny [sg:acc:m][sg:nom:m]	67		
	obrotnych [pl:acc:m][pl:gen:f]			
	[pl:gen:m]	9		
	obrotnym [pl:dat:m][pl:inst:m]	8		
	obrotnymi [pl:inst:m]	1		
opieszali	opieszalą [sg:acc:f] [sg:inst:f]	7	146	1
	opieszala [sg:nom:f]	13		
	opieszale [pl:acc:f][pl:acc:n]			
	[pl:nom:f][pl:nom:m]	33		
	opieszalego [sg:acc:m][sg:acc:n]	21		
	opieszalej [sg:gen:f]	2		
	opieszalemu [sg:dat:m]	1		
	opieszali [pl:nom:m]	6		
	opieszali [sg:acc:m][sg:nom:m]	28		
	opieszalych [pl:acc:m]			
	[pl:gen:m][pl:loc:m]	22		
	opieszalym [pl:dat:m][pl:inst:m]			
	[pl:inst:n]	10		
	opieszalymi [pl:inst:m][pl:inst:n]	3		

¹⁴⁶ English translations of the Polish target words are presented in Table A.10.

¹⁴⁷ sg, pl – singular, plural (number)

m, f, n – masculine, feminine and neuter (gender)

nom, gen, dat, acc, inst, loc, voc – nominative, genitive, dative, accusative, instrumental, locative and vocative (cases)

Base forms (adjectives)	Inflected forms	Frequency of inflected forms	Overall frequency	No on the Polish Test
osowiały	osowiałą [sg:acc:f] osowiałe [pl:acc:m][pl:nom:f][pl:nom:m] osowiałej [sg:gen:f] osowiali [pl:nom:m] osowiały [sg:nom:m] osowiałych [pl:gen:m] osowiałym [sg:inst:m]	2 12 1 6 6 5 3	35	27
pejoratywny	pejoratywny [sg:nom:m][sg:acc:m] pejoratywnego [sg:gen:m][sg:gen:n] pejoratywnemu [sg:dat:n] pejoratywnym [sg:loc:m][sg:inst:m][sg:loc:n] [sg:inst:n] pejoratywna [sg:nom:f] pejoratywną [sg:acc:f][sg:inst:f] pejoratywnej [sg:gen:f] pejoratywne [sg:nom:n][sg:acc:n][pl:nom:n] [pl:acc:n] pejoratywnych [pl: gen:m][pl:gen:f][pl:gen:n] pejoratywnymi [pl:inst:n]	25 28 1 40 7 9 3 52 8 1	174	39
skwarny	skwarną [sg:acc:f] skwarna [sg:nom:f] skwarne [pl:acc:f][pl:acc:n] skwarne [sg:gen:n] skwarnej [sg:loc:f] skwarny [sg:acc:m] skwarnych [pl:gen:m] skwarnym [sg:inst:n]	1 1 19 1 4 3 2 5	36	19
wierutny	wierutną [sg:acc:f][sg:inst:f] wierutna [sg:nom:f] wierutne [pl:acc:f][pl:nom:f] wierutnego [sg:gen:m] wierutny [sg:nom:m] wierutnym [sg:inst:n] wierutnymi [pl:inst:f]	9 30 25 1 1 8 1	75	44

APPENDIX I: OUP Quick Placement Test

The OUP Quick Placement Test (OUP, 2001) is a restricted instrument, and hence it cannot be reproduced here in its full form. In order to illustrate the content and the format of the document the first part of each section is shown below.

The test designers strongly recommend that the test be used alongside another form of assessment, as its standard error of measurement is + or – 4 points on the 60-point scale (OUP, 2001). Therefore, an additional writing task was devised and added as the last section of the test. The writing task is not part of the original OUP test.

Instructions:¹⁴⁸

This placement test will allow us to allocate you to an appropriate English language group. This test is divided into two parts:

- 1. Part One (Questions 1 – 40) – all students should respond to these.*
- 2. Part Two (Questions 41 – 60) – only the students instructed by the test supervisor should complete this section.*

When you have finished the above section(s), you will be asked to complete a short writing task.

Time: 45 minutes (30 minutes for the placement test and 15 minutes for writing)

Part 1

Questions 1 – 5

Where can you see these notices?

For questions 1 to 5, mark one letter **A**, **B** or **C** on your Answer Sheet.

1.

**Please leave
your room key
at Reception**

- A. in a shop
- B. in a hotel
- C. in a taxi

¹⁴⁸ Instructions were provided in Polish and English.

Questions 6 – 10

In this section you must choose the word which best fits each space in the text below.
For questions 6 to 10, mark one letter A, B or C on your Answer Sheet.

Scotland

(...) Some people (6) Scotland speak a different language called Gaelic. (...)

6. A on B in C at

Questions 11 – 20

In this section you must choose the word which best fits each space in the texts. For questions **11** to **20**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

Alice Guy Blaché

(...) Alice was the first to use many new inventions, (11) sound and colour.

11. **A** bringing **B** including **C** containing **D** supporting

Questions 21 – 40

In this section you must choose the word or phrase which best completes each sentence. For questions **21** to **40**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

21. The teacher encouraged her students to an English pen-friend.

- A** should write **B** write **C** wrote **D** to write

Part 2

Do not start this part unless told to do so by your test supervisor.

Questions 41 – 50

In this section you must choose the word or phrase which best fits each space in the texts.

For questions **41** to **50**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

The tallest buildings – SKYSCRAPERS

(...) A building which was many (41) high was first called a skyscraper in the United States at the end of the 19th century (...)

41. A stages B steps C storeys D levels

Questions 51 – 60

In this section you must choose the word or phrase which best completes each sentence. For questions **51** to **60**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

51. Roger’s manager to make him stay late if he hadn’t finished the work.
A insisted B warned C threatened D announced

Instructions for the writing task¹⁴⁹:

Write up to 10 sentences on ONE of the following topics.

1. *Your best friend*
2. *Your last holiday*
3. *Your everyday life in Ireland*
4. *The most scary experience in your life*

¹⁴⁹ The writing task is not part of the original OUP test.

APPENDIX J: Results of the pilot study

Table A.15. Piloting – Wordlikeness ratings and pronunciation difficulty of 40 low-wordlike nonwords

Low-wordlike nonwords	Wordlikeness ratings	Pronunciation difficulty
	1 – very unlikely to pass for a real Polish word 5 – very likely to pass for a Polish word	1 – very easy 5 – very difficult
bafroce	2.25	1.15
biznale	2.2	1.05
bostagi	1.95	1.05
bywlera	2	2
ceplira	2.1	2.1
ceptyko	1.7	1.2
dagryko	2.1	1.1
dokrapu	1.85	1.1
fezbino	2.1	2.3
fimrosy	2.35	2.2
foksela	1.9	1.15
gadlipu	2.05	1.1
gaflosy	2	1.85
gibruta	2	1.1
gubrawi	2.3	1.05
jedmuda	2.3	2.25
jestaby	1.7	1.1
jubleny	2.35	1.05
jukloty	2.2	1.05
kacmoba	1.85	1.1
kosmuca	1.7	1.25
lagroki	1.85	1.35
letrumi	1.65	1.2
lizmato	1.9	1.9
meksona	2.25	1.45
mudrygo	1.6	2.05
nazdoty	1.9	2.3
neglika	2.1	1.15
pegwika	2.35	1.2
raplido	1.3	1.05
roblewy	2	1.15
ruspale	1.7	1.25
rysundo	2.25	2.1
seprody	1.95	1.4
sopryla	2.25	1.25
toskaby	2.05	2.1
wekluga	1.9	2.2
woskane	2.35	1.25
wycmosa	1.8	2.2
zudrami	1.45	1.2
Descriptive statistics	min = 1.3 max = 2.35 μ = 1.988 SD = 0.259	min = 1.05 max = 2.3 μ = 1.48 SD = 0.467

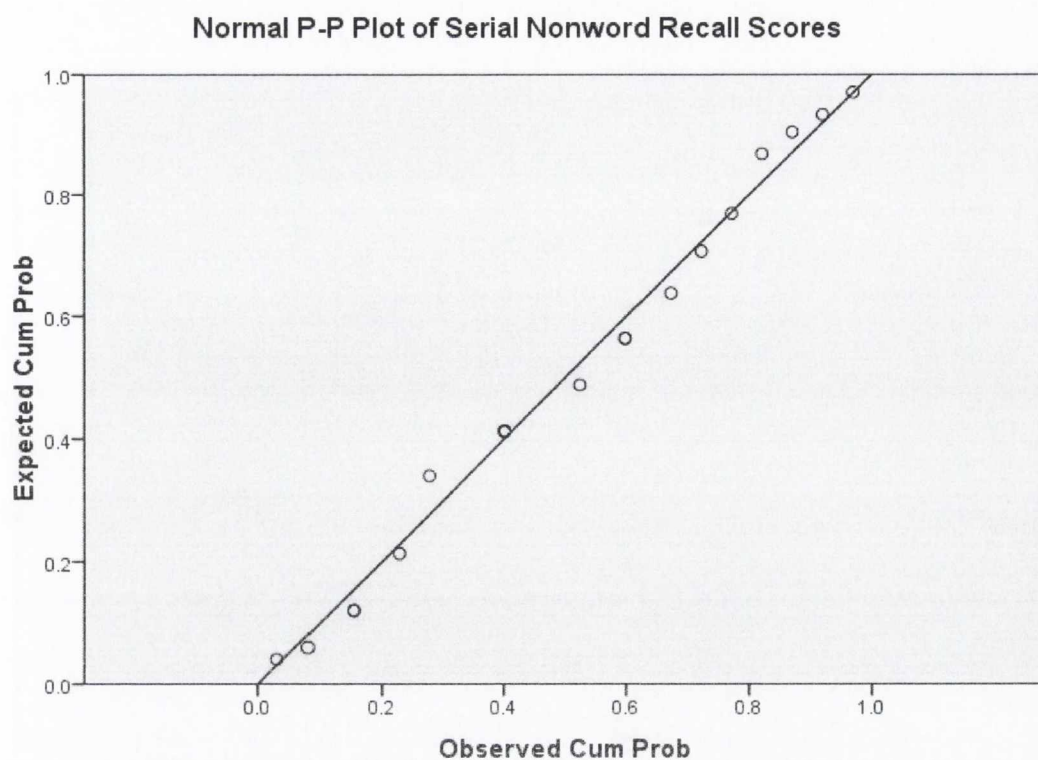


Figure A.1. Piloting – Normal probability plot of Serial Nonword Recall scores (n=20)

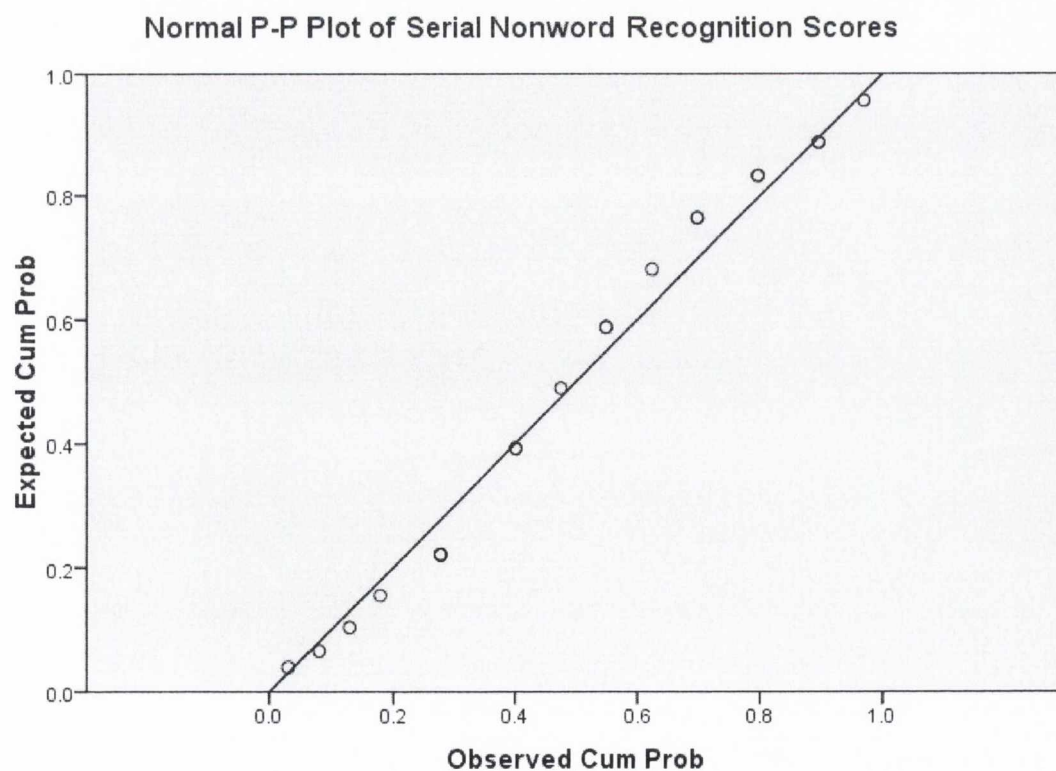


Figure A.2. Piloting – Normal probability plot of Serial Nonword Recognition scores (n=20)

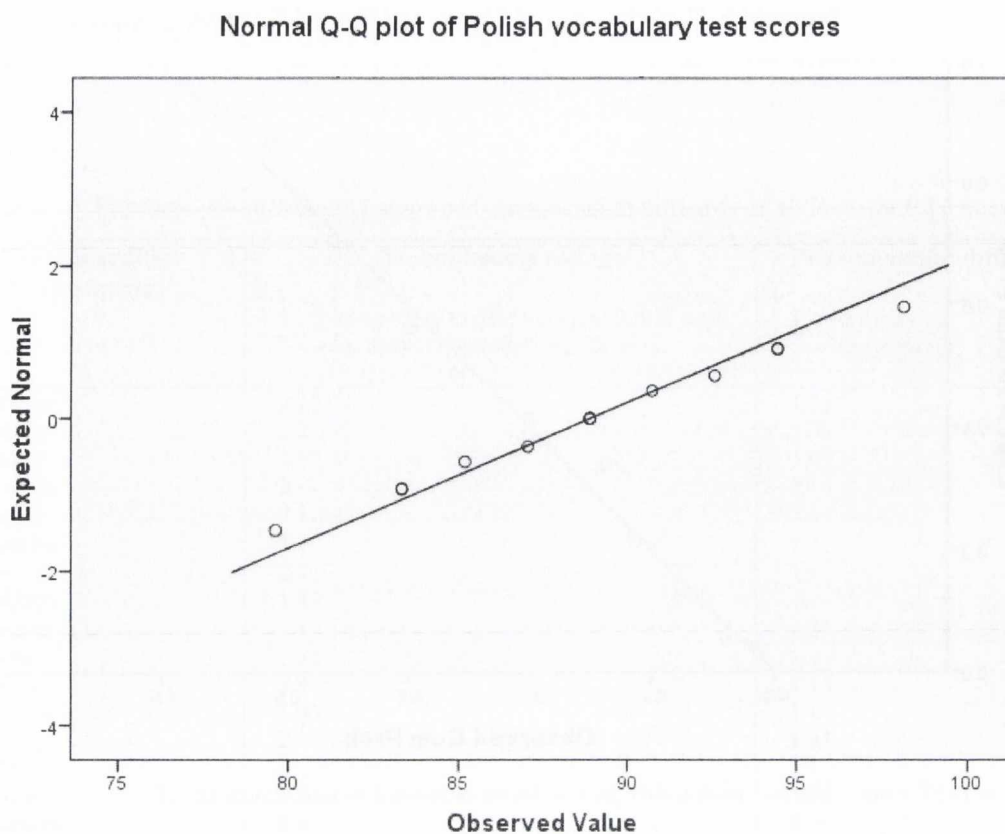


Figure A.3. Piloting – Normal probability plot of Polish Vocabulary Test scores (n=13)

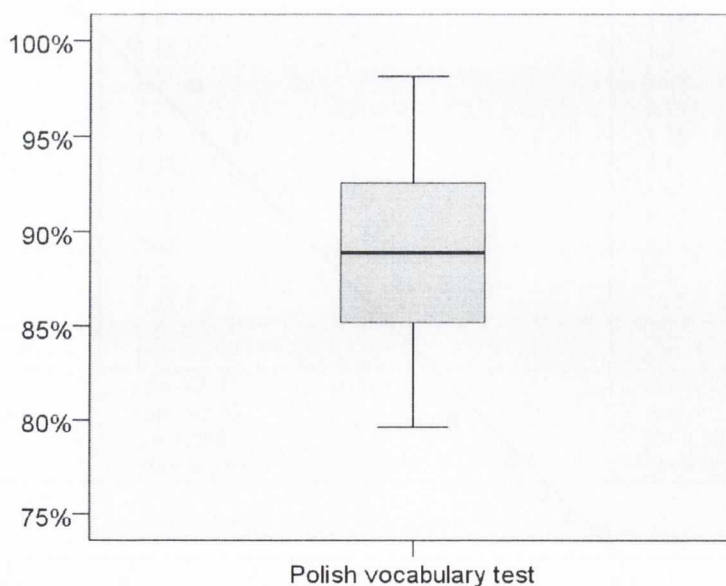


Figure A.4. Piloting - Boxplot of the distribution of Polish Vocabulary Test scores (n=13)

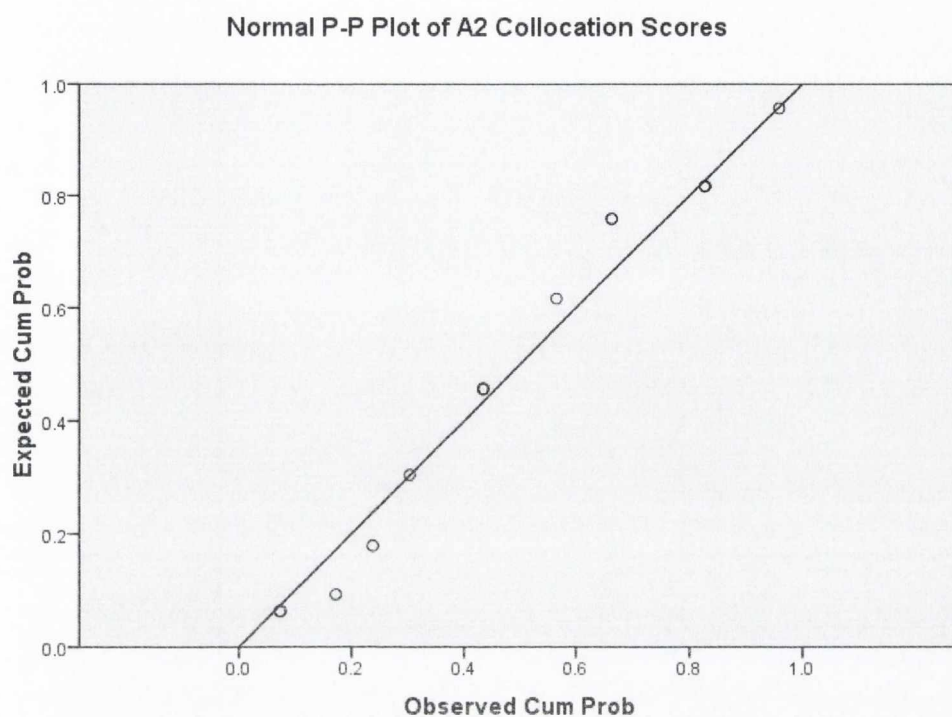


Figure A.5. Piloting – Normal probability plot of A2 Gitsaki Collocation Test scores (n=15)

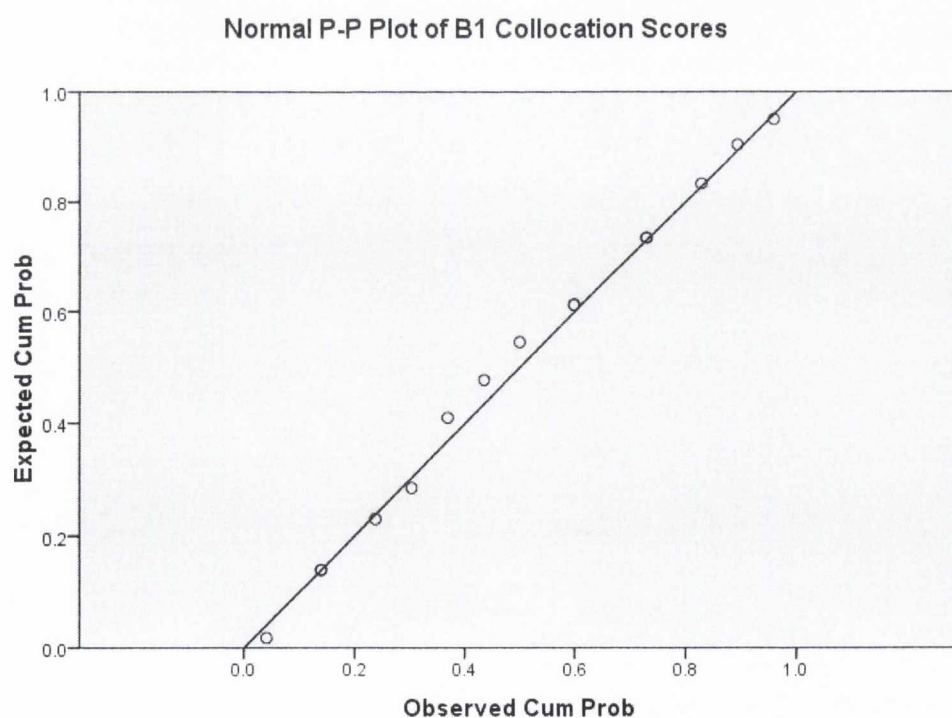


Figure A.6. Piloting – Normal probability plot of B1Gitsaki Collocation Test scores (n=15)

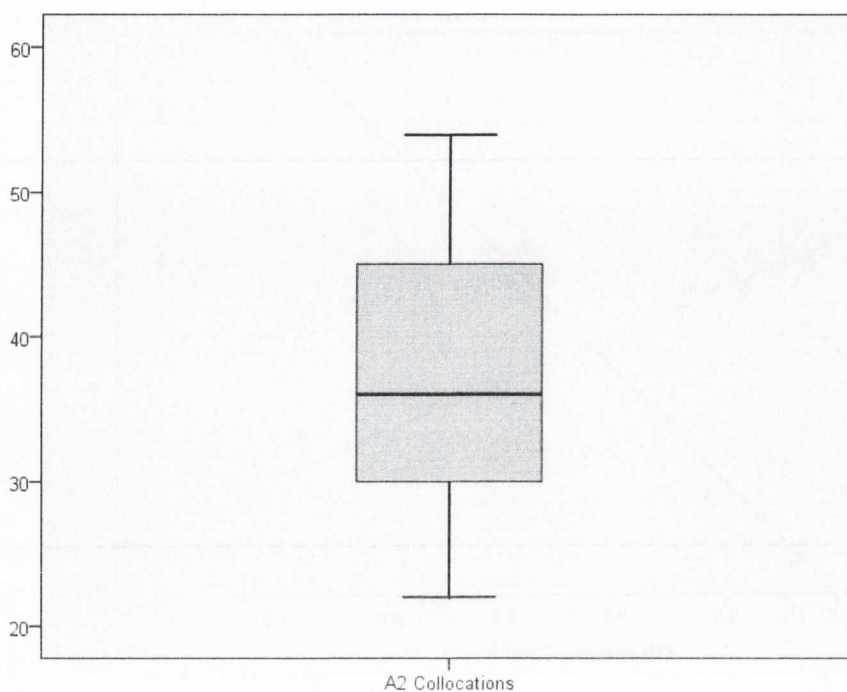


Figure A.7. Piloting - Boxplot of the distribution of A2 Gitsaki Collocation Test scores (n=15)

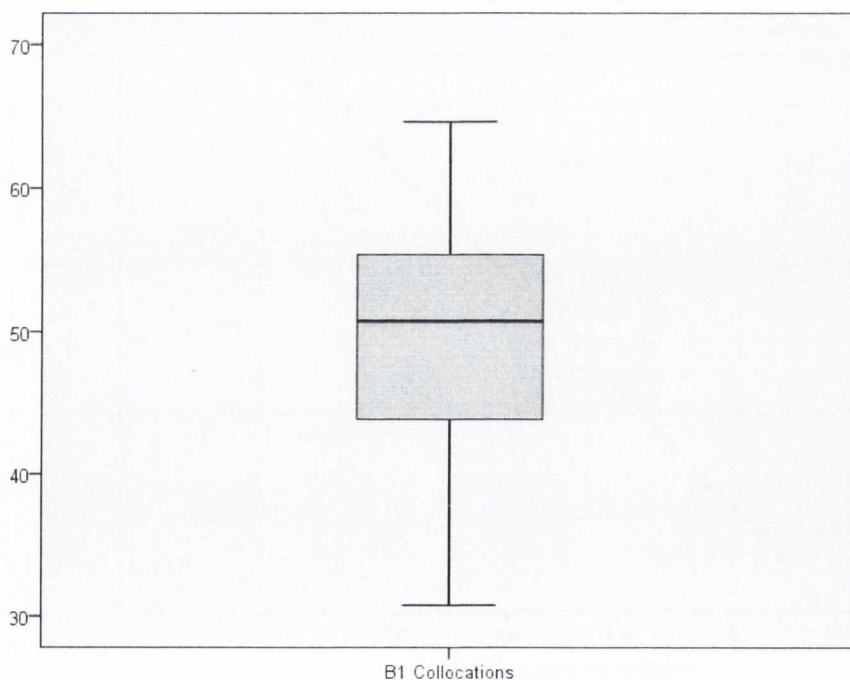


Figure A.8. Piloting - Boxplot of the distribution of B1 Gitsaki Collocation Test scores (n=15)

APPENDIX K: Results of the main study

Table A.16. Serial Nonword Recall and Recognition scores in A2 and B1 groups (Time 1)

Statistics	A2 Serial Recall	B1 Serial Recall	A2 Serial Recognition	B1 Serial Recognition
k	45	45	30	30
N	30	30	30	30
mean	15.6	17.5	p16.4	18.1
min /max	6/25	6/30	9/24	10/25
SD	4.944	6.257	3.839	3.912
Kolmogorov-Smirnov test	$D(30) = 0.096$, ns	$D(30) = 0.087$, ns	$D(30) = 0.115$, ns	$D(30) = 0.123$, ns
Shapiro-Wilk test	$D(30) = 0.979$, ns	$D(30) = 0.977$, ns	$D(30) = 0.979$, ns	$D(30) = 0.974$, ns
Skewness	-0.087, ns	-0.038, ns	0.036, ns	-0.235, ns
Kurtosis	-0.644, ns	-0.584, ns	-0.694, ns	-0.644, ns

Table A.17. Serial Nonword Recall scores in High and Low Serial Nonword Recall groups (Time 1)

Statistics	A2 Low Serial Recall group	A2 High Serial Recall group	B1 Low Serial Recall group	B1 High Serial Recall group
N	15	15	15	15
Mean	11.6	19.6	12.4	22.6
SD	2.898	2.717	3.582	3.436
Kolmogorov-Smirnov test	$D(15) = 0.219$, ns	$D(15) = 0.130$, ns	$D(15) = 0.27$, sig ¹⁵⁰	$D(15) = 0.219$, ns
Shapiro-Wilk test	$D(15) = 0.883$, ns	$D(15) = 0.953$, ns	$D(15) = 0.884$, ns	$D(15) = 0.930$, ns
Skewness	-0.717, ns	0.406, ns	-0.700, ns	.717, ns
Kurtosis	-0.848, ns	-0.623, ns	-0.846, ns	-0.252, ns
Levene's test	$F(1,28) = 0.333$, ns		$F(1,28) = 0.047$, ns	

Table A.18. Serial Nonword Recognition scores in High and Low Serial Nonword Recognition groups (Time 1)

Statistics	A2 Low Serial Recognition group	A2 High Serial Recognition group	B1 Low Serial Recognition group	B1 High Serial Recognition group
N	15	15	15	15
Mean	13.2	19.6	14.8	21.3
SD	2.052	2.197	2.416	1.981
Kolmogorov-Smirnov test	$D(15) = 0.182$, ns	$D(15) = 0.141$, ns	$D(15) = 0.189$, ns	$D(15) = 0.153$, ns
Shapiro-Wilk test	$D(15) = 0.933$, ns	$D(15) = 0.962$, ns	$D(15) = 0.913$, ns	$D(15) = 0.959$, ns
Skewness	-0.700, ns	.357, ns	-.767, ns	0.082, ns
Kurtosis	-0.017, ns	-.062, ns	-.393, ns	-0.170, ns
Levene's test	$F(1,28) = 0.072$, ns		$F(1,28) = 0.680$, ns	

¹⁵⁰ Please note that in the case of B1 low Serial Nonword Recall group the Shapiro-Wilk test statistic is insignificant ($p > 0.05$), while the Kolmogorov-Smirnov test statistic is significant ($p < 0.05$). Considering the fact that the values of skewness and kurtosis are non-significant and that the Shapiro-Wilk test (indicating no significant deviations from normality) is recommended for smaller samples like ours, we decided to use parametric tests (t-tests) rather than non-parametric tests to compare means of the B1 low Serial Recall group and the B1 high Serial Recall group.

Table A.19. Serial Nonword Recall and Recognition scores in A2 and B1 groups (Time 2)

Statistics	A2 Serial Recall	B1 Serial Recall	A2 Serial Recognition	B1 Serial Recognition
k	45	45	30	30
N	30	30	30	30
mean	15.5	17.5	16	18
min /max	6/24	7/30	8/25	9 /28
SD	4.883	6.246	4.135	4.513
Kolmogorov-Smirnov test	$D(30) = 0.092$, ns	$D(30) = .105$, ns	$D(30) = 0.104$, ns	$D(30) = 0.103$, ns
Shapiro-Wilk test	$D(30) = 0.971$, ns	$D(30) = 0.959$, ns	$D(30) = 0.981$, ns	$D(30) = 0.985$, ns
Skewness	-0.175, ns	-0.141, ns	0.232, ns	0.142, ns
Kurtosis	-0.883, ns	-0.781, ns	-0.216, ns	-0.075, ns
Levene's test	$F(1,28) = 0.938$, ns	$F(1,28) = 1.857$, ns	$F(1,28) = 1.292$, ns	$F(1,28) = 0.591$, ns

Table A.20. Polish Vocabulary Test scores in A2 and B1 groups (Time 1)

Statistics	Polish vocabulary (A2 group)	Polish vocabulary (B1 group)
N	30	30
Mean	78.1 %	81.8 %
Min /max	61.1/ 98.1 %	64.8/ 96.3 %
SD	10.703 %	9.033 %
Kolmogorov-Smirnov test	$D(30) = 0.122$, ns	$D(30) = 0.138$, ns
Shapiro-Wilk test	$D(30) = 0.943$, ns	$D(30) = 0.945$, ns
Skewness	-0.201, ns	-0.576, ns
Kurtosis	-0.938, ns	-0.142, ns
Levene's test	$F(1,28) = 2.291$, ns	$F(1,28) = 2.539$, ns

Table A.21. Exposure scores in A2 and B1 groups

Statistics	Exposure scores (A2 group)	Exposure scores (B1 group)
N	30	30
Mean	4.4 (h/d)	5.3 (h/d)
Min /max	1 / 8 (h/d)	2 / 9 (h/d)
SD	2.235 (h/d)	2.204 (h/d)
Kolmogorov-Smirnov test	$D(30) = 0.155$, ns	$D(30) = 0.125$, ns
Shapiro-Wilk test	$D(30) = 0.932$, ns	$D(30) = 0.940$, ns
Skewness	0.059, ns	0.098, ns
Kurtosis	-1.169, ns	-1.095, ns
Levene's test	$F(1,28) = 0.018$, ns	$F(1,28) = 0.108$, ns

Note: (h/d) – hours per day

Table A.22. A2 Vocabulary Levels Test scores (Time 1 and Time 2)

Statistics	A2 Vocabulary Levels Test (Time 1)	A2 Vocabulary Levels Test (Time 2)
N	30	30
Mean	49.2 %	56.1 %
Min /max	30.4/ 70.3 %	31.7/ 80.9 %
SD	12.881 %	13.296 %
Kolmogorov-Smirnov test	$D(30) = 0.130$, ns	$D(30) = 0.098$, ns
Shapiro-Wilk test	$D(30) = 0.930$, ns	$D(30) = 0.971$, ns
Skewness	0.158, ns	0.025, ns
Kurtosis	-1.337, ns	-0.966, ns
Levene's test	$F(1,28) = 0.809$, ns	$F(1,28) = 0.161$, ns

Table A.23. B1 Vocabulary Levels Test scores (Time 1 and Time 2)

Statistics	B1 Vocabulary Levels Test (Time 1)	B1 Vocabulary Levels Test (Time 2)
N	30	30
Mean	52.1 %	64.5 %
Min /max	20.3/ 76.7%	31.3/ 91.9 %
SD	16.432 %	15.164 %
Kolmogorov-Smirnov test	$D(30) = 0.129$, ns	$D(30) = 0.142$, ns
Shapiro-Wilk test	$D(30) = 0.949$, ns	$D(30) = 0.959$, ns
Skewness	-0.660, ns	-0.484, ns
Kurtosis	-0.111, ns	-0.344, ns
Levene's test	$F(1,28) = 3.645$, ns	$F(1,28) = 1.956$, ns

Table A.24. A2 and B1 Vocabulary Levels Test gain scores

Statistics	A2 Vocabulary Levels Test gain scores	B1 Vocabulary Levels Test gain scores
N	30	30
Mean	6.9 %	12.4 %
Min /max	0.7/ 16.2 %	1.1 / 26.3 %
SD	4.094 %	6.898%
Kolmogorov-Smirnov test	$D(30) = 0.090$, ns	$D(30) = 0.088$, ns
Shapiro-Wilk test	$D(30) = 0.952$, ns	$D(30) = 0.966$, ns
Skewness	0.623, ns	0.417, ns
Kurtosis	0.030, ns	-0.528, ns
Levene's test	$F(1,28) = 3.012$, ns	$F(1,28) = 2.982$, ns

Table A.25. A2 Gitsaki Collocation Test scores (Time 1 and Time 2)

Statistics	A2 Gitsaki Collocation Test (Time 1)	A2 Gitsaki Collocation Test (Time 2)
N	30	30
Mean	27.03 %	33.9 %
Min /max	4/ 54 %	6/ 62 %
SD	12.655 %	14.321 %
Kolmogorov-Smirnov test	$D(30) = 0.077$, ns	$D(30) = 0.0726$, ns
Shapiro-Wilk test	$D(30) = 0.979$, ns	$D(30) = 0.979$, ns
Skewness	0.308, ns	-0.056, ns
Kurtosis	-0.341, ns	-0.425, ns
Levene's test	$F(1,28) = 0.893$, ns	$F(1,28) = 0.263$, ns

Table A.26. B1 Gitsaki Collocation Test scores (Time 1 and Time 2)

Statistics	B1 Gitsaki Collocation Test (Time 1)	B1 Gitsaki Collocation Test (Time 2)
N	30	30
Mean	55.03 %	63.1 %
Min /max	32/ 80 %	38/ 88 %
SD	12.360 %	12.802 %
Kolmogorov-Smirnov test	$D(30) = 0.088$, ns	$D(30) = 0.082$, ns
Shapiro-Wilk test	$D(30) = 0.974$, ns	$D(30) = 0.971$, ns
Skewness	-0.093, ns	-0.052, ns
Kurtosis	-0.431, ns	-0.082, ns
Levene's test	$F(1,28) = 0.634$	$F(1,28) = 0.001$, ns

Table A.27. A2 and B1 Gitsaki Collocation Test gain scores

Statistics	A2 Gitsaki Collocation Test gain scores	B1 Gitsaki Collocation Test gain scores
N	30	30
Mean	6.9 %	8.1 %
Min /max	0 / 16 %	4/ 14 %
SD	5.241 %	2.827 %
Kolmogorov-Smirnov test	$D(30) = 0.158$, ns	$D(30) = 0.077$, ns
Shapiro-Wilk test	$D(30) = 0.950$, ns	$D(30) = 0.982$, ns
Skewness	0.306, ns	0.447, ns
Kurtosis	-1.196, ns	-0.202, ns
Levene's test	$F(1,28) = 0.757$, ns	$F(1,28) = 0.637$, ns

Table A.28. A2 and B1 Progress Vocabulary Test scores (Time 2)

Statistics	A2 Progress Vocabulary Test (Time 1)	B1 Progress Vocabulary Test (Time 2)
N	30	30
Mean	83.4 %	79.6 %
Min /max	46.6/96.6 %	53.3/96.7 %
SD	13.115	14.033
Kolmogorov-Smirnov test	$D(30) = 0.197$, sig	$D(30) = 0.159$, sig
Shapiro-Wilk test	$D(30) = 0.877$, sig	$D(30) = 0.889$, sig
Skewness	-1.209, sig	-0.652, ns
Kurtosis	0.546, ns	-0.902, ns
Levene's test	$F(1,28) = 10.578$, sig	$F(1,28) = 33.165$, sig

Table A.29. A2 and B1 Progress Collocation Test scores (Time 2)

Statistics	A2 Progress Collocation Test (Time 1)	B1 Progress Collocation Test (Time 2)
N	30	30
Mean	58.5 %	68.1 %
Min /max	46.6/78.8 %	47.5/90.4 %
SD	11.180 %	14.949
Kolmogorov-Smirnov test	$D(30) = 0.163$, sig	$D(30) = 0.153$, ns
Shapiro-Wilk test	$D(30) = 0.867$, sig	$D(30) = 0.069$, sig
Skewness	0.573, ns	-0.105, ns
Kurtosis	-1.033, ns	-1.58, ns
Levene's test	$F(1,28) = 10.739$, sig	$F(1,28) = 0.926$, ns

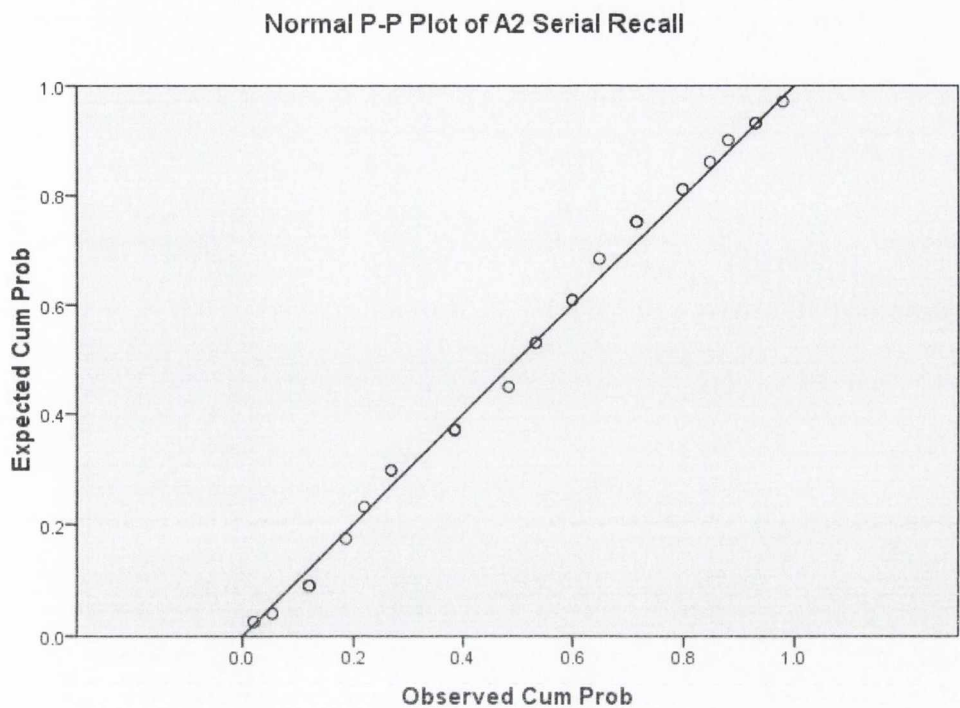


Figure A.9. Normal probability plot of A2 Serial Nonword Recall scores at Time 1(n=30)

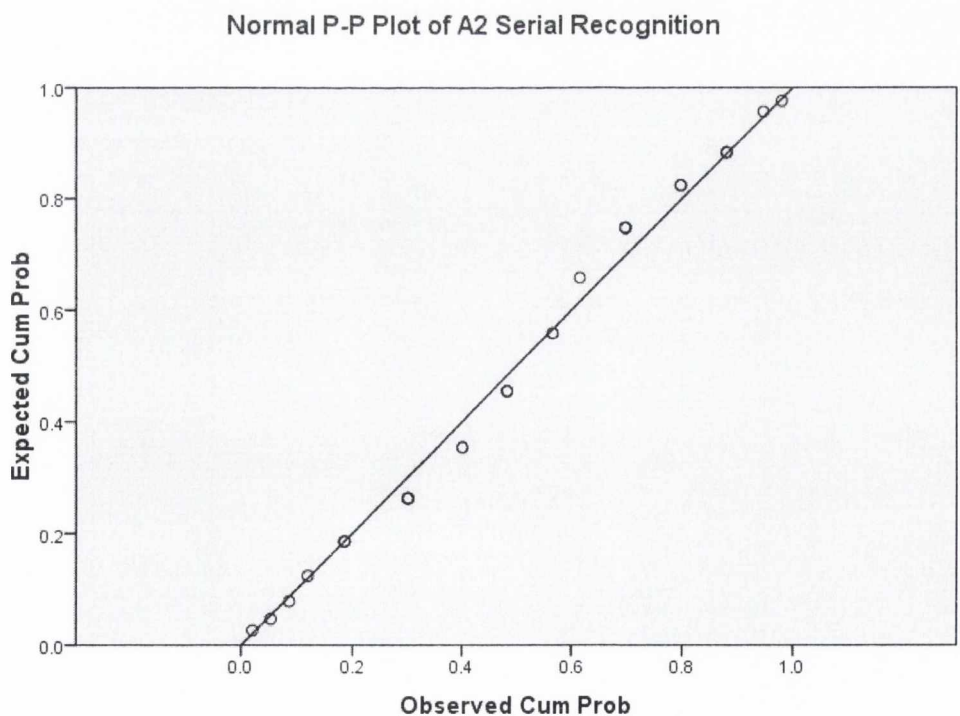


Figure A.10. Normal probability plot of A2 Serial Nonword Recognition scores at Time 1 (n=30)

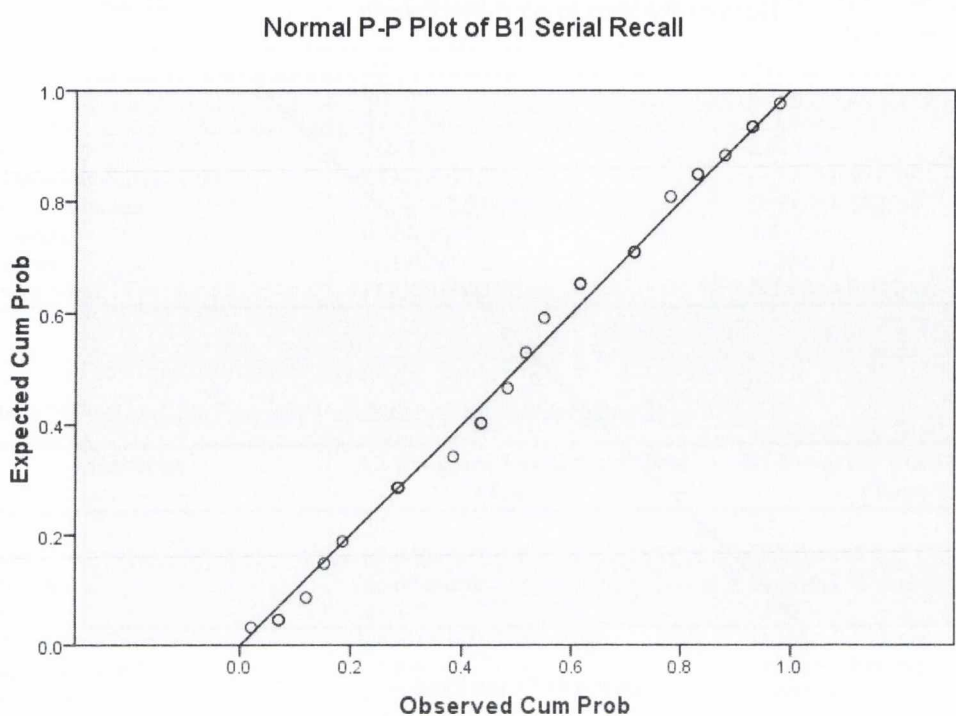


Figure A.11. Normal probability plot of B1 Serial Nonword Recall scores at Time 1 (n=30)

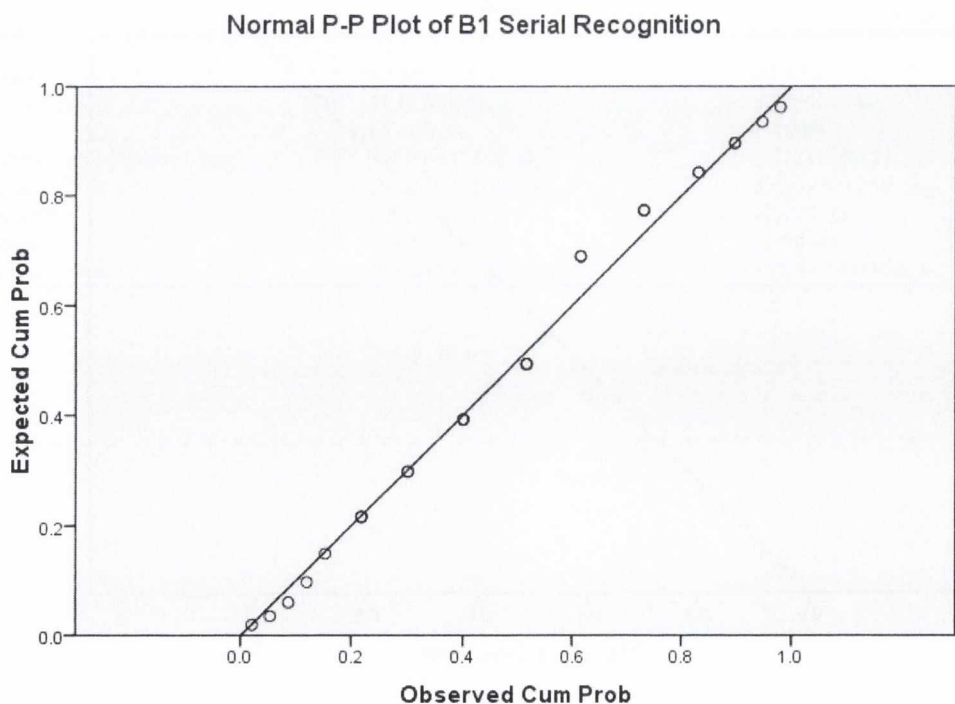


Figure A.12. Normal probability plot of B1 Serial Nonword Recognition scores at Time 1 (n=30)

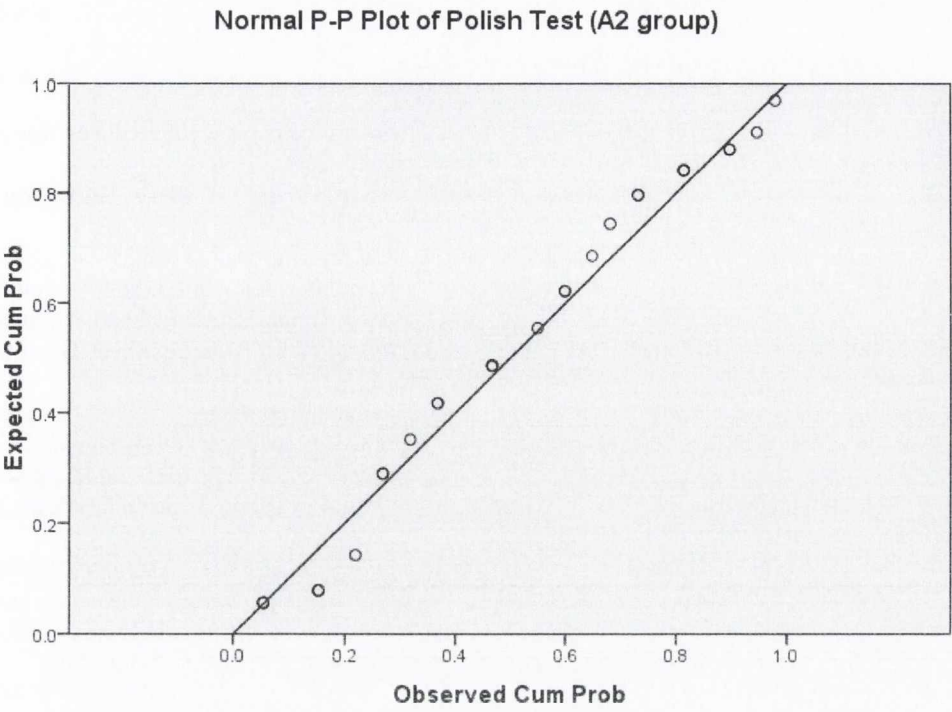


Figure A.16. Normal probability plot of A2 Polish Vocabulary Test scores at Time 1 (n=30)

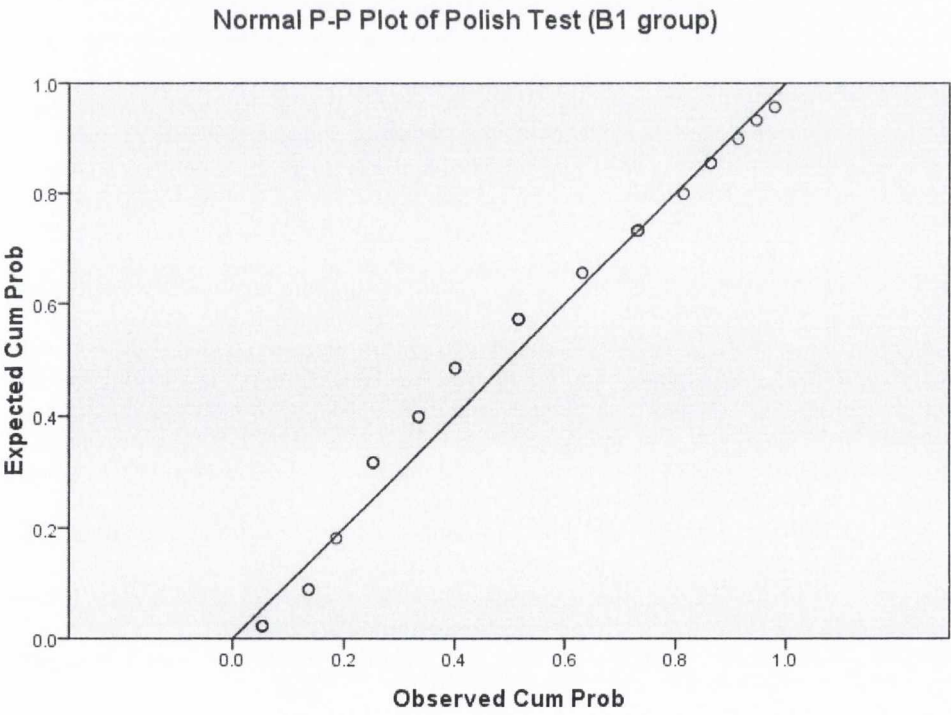


Figure A.17. Normal probability plot of B1 Polish Vocabulary Test scores at Time 1 (n=30)

References

- Adamopoulou, R. (2000). The equivalence of a computer-delivered and a paper-delivered Yes/No Vocabulary Test. M.A. thesis, University of Wales, Swansea.
- Al-Hazemi, H. (1993). Low level EFL vocabulary tests for Arabic speakers. M.A. thesis, University of Wales, Swansea.
- Anderson, R. C. & Freebody, Peter (1981). Vocabulary knowledge. In Guthrie, J. T. (Ed.), *Comprehension and Teaching: Research Reviews*. Newark: International Reading Association.
- Anderson, R. C. & Freebody, Peter (1983). Reading comprehension and the assessment and acquisition of word knowledge. In Huston, B. (Ed.), *Advances in Reading/Language Research*. Greenwich: JAI Press.
- Andrade, Jackie (2001). *Working Memory in Perspective*. Hove: Psychology Press.
- Archibald, Lisa M. D. & Gathercole, Susan Elizabeth (2007). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics*, 28(4): 587-606.
- Atkins, Paul W. B. & Baddeley, Alan D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19(4): 537-552.
- Atkinson, Richard C. & Shiffrin, Richard M. (1968). Human memory: A proposed system and its control processes. In Spence, K. W. & Spence, J. T. (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 2). New York: Academic Press.
- Bachman, Lyle F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Baddeley, Alan D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, 18A(4): 362-365.
- Baddeley, Alan D. (1968). How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology*, 20A(3): 249-264.
- Baddeley, Alan D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 85(3): 139-152.
- Baddeley, Alan D. (1986). *Working Memory*. Oxford: Oxford University Press.
- Baddeley, Alan D. (1990). *Human Memory: Theory and Practice*. Hove: Erlbaum.
- Baddeley, Alan D. (1992). Is working memory working? The Fifteenth Bartlett Lecture. *The Quarterly Journal of Experimental Psychology*, 44A(1): 1-31.
- Baddeley, Alan D. (1993). Short-term phonological memory and long-term learning: A Case Study. *European Journal of Cognitive Psychology*, 5: 129-148.
- Baddeley, Alan D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11): 417-423.
- Baddeley, Alan D. (2003a). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3): 85-107.
- Baddeley, Alan D. (2003b). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4: 829-839.
- Baddeley, Alan D. (2007). *Working Memory, Thought, and Action*. Oxford: Oxford University Press.
- Baddeley, Alan D., Bressi, S., Della Sala, Sergio, Logie, Robert H. & Spinnler, H. (1991). The decline of working memory in Alzheimer's Disease: A longitudinal study. *Brain*, 114: 2521-2542.
- Baddeley, Alan D., Chincotta, Dino, Stafford, Lorenzo & Turk, David (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *The Quarterly Journal of Experimental Psychology*, 55A(2): 353-369.
- Baddeley, Alan D., Gathercole, Susan Elizabeth & Papagno, Costanza (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1): 158-173.
- Baddeley, Alan D. & Hitch, Graham J. (1974). Working memory. In Bower, G. (Ed.), *The Psychology of Learning and Motivation*. New York: Academic Press. 47-90.
- Baddeley, Alan D., Lewis, Vivien & Vallar, Giuseppe (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology*, 36A(2): 233-252.
- Baddeley, Alan D. & Logie, Robert H. (1992). Auditory imagery and working memory. In Reisberg, D. (Ed.), *Auditory Imagery*. Hillsdale: Erlbaum.
- Baddeley, Alan D., Logie, Robert H., Bressi, S., Della Sala, Sergio & Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology*, 38A(4): 603-618.
- Baddeley, Alan D., Papagno, Costanza & Vallar, Giuseppe (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27(5): 586-595.
- Baddeley, Alan D., Thomson, Neil & Buchanan, Mary (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14: 575-589.

- Baddeley, Alan D. & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, 24: 490-502.
- Bahns, Jens & Eldaw, Moira (1993). Should we teach EFL students collocations? *System*, 21(1): 101-114.
- Bailey, Todd M. & Hahn, Ulrike (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44: 568-591.
- Barlow, Michael & Kemmer, Suzanne (1999). *Usage-based Models of Language* Cambridge: Cambridge University Press.
- Barrell, Ray, FitzGerald, John & Riley, Rebecca (2007). EU enlargement and migration: Assessing the macroeconomic impacts. NIESR Discussion Paper No. 292.
- Bartsch, Sabine (2004). *Structural and Functional Properties of Collocations in English: A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence*. Tübingen: Gunter Narr Verlag Tübingen.
- Bauer, Laurie & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6(4): 253-279.
- Beeckmans, Renaud, Eyckmans, June, Janssens, Vera & Velde, Hans Van de (2001). Examining the Yes/No Vocabulary Test: Some methodological issues in theory and practice. *Language Testing*, 18(3): 235-274.
- Beglar, David (2009). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 26(4): 1-22.
- Beglar, David & Hunt, Alan (1999). Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests. *Language Testing*, 16(2): 131-162.
- Belleville, Sylvie, Peretz, Isabelle & Arguin, Martin (1992). Contribution of articulatory rehearsal to short-term memory: Evidence from a case of selective disruption. *Brain and Language*, 43(4): 713-746.
- Benson, Morton, Benson, Evelyn & Ilson, Robert (1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins Publishing Company.
- Benson, Morton, Benson, Evelyn & Ilson, Robert (1997). *The BBI Dictionary of English Word Combinations*. Poznań: OMNIBUS.
- Bertram, Raymond, Baayen, R. Harald & Schreuder, Robert (2000). Effects of family size for complex words. *Journal of Memory and Language*, 42(3): 390-405.
- Bertram, Raymond, Laine, Matti & Virkkala, Minna Maria (2000). The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. *Scandinavian Journal of Psychology*, 41(4): 287-296.
- Biber, Douglas, Conrad, Susan & Reppen, Randi (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward (1999). Lexical expressions in speech and writing. *Longman Grammar of Spoken and Written Language*. Harlow: Pearson Education.
- Bielska, Joanna (2002). On the relationship between psychological type, language learning strategies and second language achievement. Ph.D. thesis, University of Silesia.
- Bishop, Dorothy V. M. (2006). Beyond words: Phonological short-term memory and syntactic impairment in specific language impairment. *Applied Psycholinguistics*, 27(4): 545-547
- Biskup, Danuta (1992). L1 influence on learners' renderings of English collocations: A Polish-German empirical study. In Arnaud, P. J. L. & Béjoint, H. (Eds.), *Vocabulary and Applied Linguistics*. London: Macmillan.
- Boeri, Tito & Brücker, Herbert (2005). Migration, xo-ordination failures and EU enlargement. IZA discussion paper no. 1600.
- Bolander, Maria (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In Hyltenstam, K. & Obler, L. K. (Eds.), *Bilingualism across the Lifespan*. Cambridge: Cambridge University Press. 73-86.
- Bonk, William J. (2001). Testing ESL learners' knowledge of collocations. In Hudson, T. & Brown, J. D. (Eds.), *A Focus on Language Test Development: Expanding the Language Proficiency Construct Across a Variety of Tests*. Honolulu: University of Hawai'i Press.
- Bowey, Judith A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study *Applied Psycholinguistics*, 22(3): 441-469.
- Broadbent, Donald Eric (1958). *Perception and Communication*. London: Pergamon Press.
- Brown, Roger (1973). *A First Language: The Early Stages*. Cambridge: Harvard University Press.
- Buchsbaum, Bradley R. & D'Esposito, Mark (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience*, 20(5): 762-778.

- Bullock, Daniel (2004). Adaptive neural models of queuing and timing in fluent action *Trends in Cognitive Sciences*, 8(9): 426-433.
- Burgess, Neil & Hitch, Graham J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, 31(4): 429-460.
- Burgess, Neil & Hitch, Graham J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106(3): 551-581.
- Burgess, Neil & Hitch, Graham J. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences*, 9(11): 535-541.
- Burgess, Neil & Hitch, Graham J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55(4): 627-652.
- Burmeister, Hartmut & Ufert, Detlef (1980). Strategy switching. In Felix, S. (Ed.), *Second Language Development*. Tübingen: Gunter Narr.
- Cameron-Faulkner, Thea, Lieven, Elena V.M. & Tomasello, Michael (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27(6): 843-873.
- Cameron, Lynne (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, 6(2): 145-173.
- Campbell, Donald T. (1988). *Methodology and Epistemology for Social Science: Selected Papers*. London: University of Chicago Press.
- Campbell, Donald T. & Stanley, Julian C. (1963). *Experimental and Quasi-experimental Designs for Research* Boston: Houghton Mifflin Company.
- Campion, Mary E. & Elley, Warwick B. (1971). *The Academic Vocabulary List*. Wellington: New Zealand Council for Educational Research.
- Cheung, Him (1996). Nonword span as a unique predictor of second language vocabulary learning. *Developmental Psychology*, 32(5): 867-873.
- Chujo, Kiyomi (2004). Measuring vocabulary levels of English textbooks and tests using a BNC lemmatised high frequency word list. *Language and Computers*, 19: 231-249.
- Chun, Dorothy M. & Payne, J. Scott (2004). What makes students click: Working memory and look-up behavior. *System*, 32(4): 481-503.
- Church, Kenneth Ward, Hanks, Patrick & Moon, Rosamund (1994). Lexical substitutability. In Atkins, P. W. B. & Zampolli, A. (Eds.), *Computational Approaches to the Lexicon*. Oxford: Oxford University Press.
- Cieślicka, Anna (2000). The effect of language proficiency and L2 vocabulary learning strategies on patterns of bilingual lexical processing. *Poznań Studies in Contemporary Linguistics*, 36: 27-53.
- Clare, Antonia & Wilson, J. J. (2006). *Total English: Intermediate*. Harlow: Longman.
- Cobb, Tom (2000). One size fits all? Francophone learners and English vocabulary tests. *Canadian Modern Language Review*, 57(2): 295-324.
- Cobb, Tom & Horst, Marlise (2004). Is there a room for an AWL in French? In Laufer, B. & Bogaards, P. (Eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing*. Amsterdam: John Benjamins.
- Cocchini, Gianna, Logie, Robert H., Della Sala, Sergio, MacPherson, Sarah E. & Baddeley, Alan D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, 30(7): 1086-1095.
- Code, Chris (1994). Speech automatism production in aphasia. *Journal of Neurolinguistics*, 8(2): 135-148.
- Cohen, Gillian, Eysenck, Michael W. & Le Voi, Martin E. (1986). *Memory: A Cognitive Approach*. Milton Keynes: Open University Press.
- Conklin, Kathy & Schmitt, Norbert (2007). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?
- Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 4: 161-169.
- Conrad, R. & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, 55: 429-432.
- Cook, Vivian (2003). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Cowie, Anthony P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2: 223-235.
- Cowie, Anthony P. (1994). Phraseology. In Asher, R. E. & Simpson, J. M. Y. (Eds.), *The Encyclopedia of Language and Linguistics (Volume 8)*. Oxford: Pergamon.

- Cowie, Anthony P. (1998a). Introduction. In Cowie, A. P. (Ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Cowie, Anthony P. (1998b). Phraseological dictionaries: Some east-west comparisons. In Cowie, A. P. (Ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Coxhead, Averil (2000). A New Academic Word List *TESOL Quarterly*, 34(2): 213-238.
- Coxhead, Averil (2002). The academic word list: A corpus-based word list for academic purposes. In Ketteman, B. & Marks, G. (Eds.), *Teaching and Language Corpora (TALC) 2000 Conference Proceedings*. Atlanta: Rodopi.
- Craik, Fergus I. M. & Lockhart, Robert S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6): 671-684.
- Craik, Fergus I. M. & Watkins, Michael J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 12(6): 599-607.
- Crano, William D., Kenny, David A. & Campbell, Donald T. (1972). Does intelligence cause achievement: A cross-lagged panel analysis. *Journal of Educational Psychology*, 63(3): 258-275.
- Crano, William D. & Mellon, Phyllis M. (1978). Causal influence of teachers' expectations on children's academic performance: A cross-lagged panel analysis. *Journal of Educational Psychology*, 70(1): 39-49.
- Crowdy, Steve (1993). Spoken corpus design *Literary and Linguistic Computing*, 8(4): 259-265.
- Daller, Helmut, Milton, James & Treffers-Daller, Jeanine (2007). Editor's introduction: conventions, terminology and an overview of the book. In Daller, H. M., Milton, J. & Treffers-Daller, J. (Eds.), *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- De Groot, Annette M. B. (1995). Determinans of bilingual lexiosemanic organisation. *Computer Assisted Language Learning*, 8: 151-180.
- De Villiers, Jill G. (1985). Learning how to use verbs: Lexical coding and the influence of the input. *Journal of Child Language*, 12(3): 587-595.
- Dębowski, Łukasz, Hajnicz, Elżbieta, Przepiórkowski, Adam & Woliński, Marcin (2006). IPI PAN (2nd edition). Retrieved October 2, 2007, from <http://korpus.pl/>.
- Della Sala, Sergio, Baddeley, Alan D., Papagno, Costanza & Spinnler, H. (1995). Dual-task paradigm: A means to examine the central executive. In Grafman, J., Holyoak, K. J. & Boller, F. (Eds.), *Structure and Functions of the Human Prefrontal Cortex*. New York: New York Academy of Sciences.
- Dörnyei, Zoltán, Durow, Valerie & Zahran, Khawla (2004). Individual differences and their effects on formulaic sequence acquisition. In Schmitt, N. (Ed.), *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins Publishing Company.
- Drinkwater, Stephen, Eade, John & Garapich, Michal (2006). Poles apart? EU enlargement and the labour market outcomes of immigrants in the UK. IZA Discussion Paper No. 2410.
- Duff, Simon C. & Logie, Robert H. (2001). Processing and storage in working memory span. *The Quarterly Journal of Experimental Psychology*, 54A(1): 31-48.
- Durrant, Philip (2008). High frequency collocations and second language learning. Ph.D. thesis, University of Nottingham.
- Durrant, Philip & Schmitt, Norbert (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47: 157-177.
- Ebbinghaus, Hermann (1885). *Memory: A Contribution to Experimental Psychology*. Translated by Henry A. Ruger & Clara E. Bussenius (1913). Originally published in New York by Teachers College, Columbia University.
- Edwards, Jan, Beckman, Mary E. & Munson, Benjamin (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language and Hearing Research*, 47(2): 421-436.
- Ellis, Nick C. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition*, 18(1): 91-126.
- Ellis, Nick C. (2001). Memory for Language. In Robinson, P. (Ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.
- Ellis, Nick C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2): 143-188.
- Ellis, Nick C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In Doughty, C. J. & Long, M. H. (Eds.), *The Handbook of Second Language Acquisition*. Oxford: Blackwell.

- Ellis, Nick C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2): 305-352.
- Ellis, Nick C. & Beaton, Alan (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology*, 46A: 533-558.
- Ellis, Nick C. & Schmidt, Richard (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, 19: 145-171.
- Ellis, Nick C. & Sinclair, Susan G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology*, 49A(1): 234-250.
- Ellis, Rod & Barkhuizen, Gary (2005). *Analysing Learner Language*. Oxford: Oxford University Press.
- Engels, L.K. (1968). The fallacy of word counts. *International Review of Applied Linguistics in Language Teaching*, 6(3): 213-231.
- Evans, Virginia & Dooley, Jenny (1999). *Enterprise 3: Grammar*. Newbury: Express Publishing.
- Eyckmans, June (2004). *Measuring Receptive Vocabulary Size*. Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap).
- Eyckmans, June (2009). Toward an assessment of learners' receptive and productive syntagmatic knowledge. In Bartfield, A. & Gyllstad, H. (Eds.), *Researching Collocations in Another Language: Multiple Interpretations*. London: Palgrave MacMillan.
- Eyckmans, June, Boers, Frank & Stengers, Helene (2006). The discriminating collocations test: A corpus-based measure of phrasal knowledge. Paper presented at the EUROSLA 16, Antalya, Turkey, 13-16 September 2006.
- Eyckmans, June, Velde, Hans Van de & van Hout, Roeland (2007). Learners' response behaviour in Yes/No Vocabulary Tests. In Daller, H., Milton, J. & Treffers-Daller, J. (Eds.), *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- Fan, May (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELJ Journal*, 31: 105-119.
- Farghal, Mohammed & Obiedat, Hussein (1995). Collocations: A neglected variable in EFL. *IRAL*, 33(4): 315-331.
- Farmer, Eric W., Berman, Jonathan V. F. & Fletcher, Yvonne L. (1986). Evidence for a visuo-spatial scratch-pad in working memory *The Quarterly Journal of Experimental Psychology*, 38(4): 675-688.
- Field, Andy P. (2005). *Discovering Statistics Using SPSS*. London: SAGE.
- Firth, John Rupert (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Firth, John Rupert (1974). A synopsis of linguistic theory, 1930-55. *Studies in Linguistic Awareness*. Oxford: Basil Blackwell.
- Fitzpatrick, Tess (2007). Productive vocabulary tests and the search for concurrent validity. In Daller, H. M., Milton, J. & Treffers-Daller, J. (Eds.), *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- Flege, J. E. & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26(1-34).
- Foster, Pauline (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In Bygate, M., Skehan, P. & Swain, M. (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. London: Longman.
- Fouser, Joseph (2000). Sociolinguistic transfer in Japanese as a third language. Ph.D. thesis, Trinity College Dublin.
- French, Leif M. (2003). Phonological working memory and L2 acquisition: A developmental study of Quebec Francophone children learning English. Ph.D. thesis, Université Laval.
- French, Leif M. (2004). Phonological working memory and L2 acquisition: A developmental study of Quebec Francophone children learning English. *Dissertation Abstracts International, A: The Humanities and Social Sciences*, 65(2): 487.
- French, Leif M. & O'Brien, Irena Socha (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29: 463-487.
- Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, 5(2): 253-286.
- Gairns, Ruth & Redman, Stuart (1986). *Working with Words: A Guide to Teaching and Learning Vocabulary*. Cambridge: Cambridge University Press.
- Gathercole, Susan Elizabeth (1992). The nature and uses of working memory. In Morris, P. E. & Gruneberg, M. (Eds.), *Theoretical Aspects of Memory*. London: Routledge.

- Gathercole, Susan Elizabeth (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, 23(1): 83-94.
- Gathercole, Susan Elizabeth (2006a). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27(4): 513-543.
- Gathercole, Susan Elizabeth & Baddeley, Alan D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, 81: 439-454.
- Gathercole, Susan Elizabeth, Frankish, Clive R., Pickering, Susan J. & Peaker, Sarah M. (1999a). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1): 84-95.
- Gathercole, Susan Elizabeth & Hitch, Graham J. (1993). Developmental changes in short-term memory. In Collins, A., Gathercole, S. E., Conway, M. A. & Morris, P. E. (Eds.), *Theories of Memory*. Hove: Lawrence Erlbaum Associates.
- Gathercole, Susan Elizabeth, Hitch, Graham J., Service, Elisabeth & Martin, Amanda J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, 33(6): 966-979.
- Gathercole, Susan Elizabeth & Martin, Amanda J. (1996). Interactive processes in phonological memory. In Gathercole, S. E. (Ed.), *Models of Short-term Memory*. Hove, England: Psychology Press. 73-100.
- Gathercole, Susan Elizabeth & Pickering, Susan J. (1999). Estimating the capacity of the phonological short-term store. *International Journal of Psychology*, 34(5 & 6): 378-382.
- Gathercole, Susan Elizabeth, Pickering, Susan J., Hall, Melanie & Peaker, Sarah M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology*, 54A(1): 1-30.
- Gathercole, Susan Elizabeth, Service, Elisabeth, Hitch, Graham J., Adams, Anne-Marie & Martin, Amanda J. (1999b). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13(1): 65-77.
- Gathercole, Susan Elizabeth & Thorn, Annabel S. C. (1998). Phonological short-term memory and foreign language learning. In Healy, A. F. & Bourne, L. E. (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*. London: Lawrence Erlbaum Associates.
- Gathercole, Susan Elizabeth, Willis, C., Emslie, Hazel & Baddeley, Alan D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12(349-367).
- Gathercole, Susan Elizabeth, Willis, Catherine S., Emslie, Hazel & Baddeley, Alan D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28(5): 887-898.
- Gitsaki, Christina (1999). *Second Language Lexical Acquisition: A Study of the Development of Collocational Knowledge*. London: International Scholars Publications.
- Gough, Chris (2000). *English Vocabulary Organizer: 100 topics for self-study*. Hove: Language Teaching Publications Series.
- Goulden, Robin, Nation, I. S. P. & Read, John (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11: 341-363.
- Graham, Athelia (2008). The effects of homography on computer-generated high frequency word lists. M.A. thesis, Brigham Young University.
- Granger, Sylviane (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Cowie, A. P. (Ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press. 145-160.
- Greene, John D. W., Hodges, John R. & Baddeley, Alan D. (1995). Autobiographical memory and executive functions in early dementia of Alzheimer type. *Neuropsychologia*, 33: 1647-1670.
- Gregg, Vernon H., Freedman, Carmen M. & Smith, Deana K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology*, 80(3): 363-374.
- Grossberg, Stephen (1978). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 17(3): 199-219.
- Grossberg, Stephen (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, 11(1): 23-63.
- Gupta, Prahlad (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology*, 56A(7): 1213-1236.
- Gupta, Prahlad (2005). Primacy and recency in nonword repetition. *Memory*, 13(3/4): 318-324.

- Gupta, Prahlad, Lipinski, John & Aktunc, Emrah (2005). Re-examining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cueing, and item recall. *Memory & Cognition*, 33(6): 1001-1016.
- Gupta, Prahlad & MacWhinney, Brian (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, 59(2): 267-333.
- Gupta, Prahlad & Tisdale, Jamie (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, 61: 481-502.
- Gyllstad, Henrik (2005). Words that go together well: Developing test formats for measuring learner knowledge of English collocations. *The Department of English in Lund: Working Papers in Linguistics*, 5: 1-39.
- Gyllstad, Henrik (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In Bartfield, A. & Gyllstad, H. (Eds.), *Researching Collocations in Another Language: Multiple Interpretations*. London: Palgrave MacMillan.
- Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word*, 27(3): 241-292.
- Halliday, M. A. K. (1966). Lexis as a linguistic level. In Bazell, C. E., Catford, J. C., Halliday, M. A. K. & Robins, R. H. (Eds.), *In Memory of J. R. Firth*. London: Longmans.
- Hartley, Tom & Houghton, George (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35(1): 1-31.
- Hay, Jennifer, Pierrehumbert, Janet & Beckman, Mary E. (2003). Speech perception, well-formedness and the statistics of the lexicon. In Local, J., Ogden, R. & Temple, R. A. M. (Eds.), *Phonetic Interpretation*. Cambridge: Cambridge University Press.
- Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In Delafresnaye, J. F. (Ed.), *Brain Mechanisms and Learning*. New York: Oxford University Press.
- Hickey, Francesca & Kuiper, Koenraad (2000). A deep depression covers the South Tasman Sea: New Zealand Meteorological Office weather. In Bell, A. & Kuiper, K. (Eds.), *New Zealand English*. Amsterdam: John Benjamins.
- Hitch, Graham J., Flude, Brenda & Burgess, Neil (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language*, 61(1): 97-111.
- Hoey, Michael (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Hoey, Michael (2005). *Lexical Priming*. London: Routledge Taylor & Francis Group.
- Hoff, Erika & Naigles, Letitia G. (2002). How children use input to acquire a lexicon. *Child Development*, 73(2): 418-433.
- Hoffmann, Sebastian & Evert, Stefan (2007). BNCweb (CQP edition). Retrieved December 2, 2007, from <http://bncweb.lancs.ac.uk/>.
- Hoffmann, Sebastian, Evert, Stefan, Smith, Nicholas, Lee, David & Prytz, Ylva Berglund (2008). *Corpus Linguistics with BNCweb – A Practical Guide*. Oxford: Peter Lang.
- Horgan, Dianne (1980). Nouns: Love 'em or leave 'em. In Teller, V. & White, S. J. (Eds.), *Studies in Child Language and Multilingualism*. New York: Annals of the New York Academy of Sciences.
- Hornby, Albert Sydney (2005). *Oxford Advanced Learner's Dictionary: 7th edition*. Oxford: Oxford University Press.
- Horst, Marlise & Meara, Paul (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2): 207-223.
- Houghton, George (1990). The problem of serial order: A neural network model of sequence learning and recall. In Dale, R., Mellish, C. & Zock, M. (Eds.), *Current Research in Natural Language Generation*. London: Academic Press.
- Howard, D. & Franklin, S. (1990). Memory without rehearsal. In Vallar, G. & Shallice, T. (Eds.), *Neuropsychological Impairments of Short-term Memory*. Cambridge: Cambridge University Press.
- Howarth, Peter (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1): 24-44.
- Huibregtse, Ineke, Admiraal, Wilfried & Meara, Paul (2002). Scores on a Yes/No vocabulary test: Correction for guessing and response style. *Language Testing*, 19(3): 227-245.
- Hulme, Charles, Maughan, Sarah & Brown, Gordon D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30(6): 685-701.
- Hulme, Charles, Thomson, Neil, Muir, Clare & Lawrence, Amanda (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, 38(2): 241-253.

- Hulstijn, Jan H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91(4): 663-667.
- Hunston, Susan (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Jafarpur, Abdoljavad (1999). What's magical about the rule-of-two for constructing C-tests? *RELC Journal*, 30: 86-100.
- James, William (1890/1931). *The Principles of Psychology, Volume 1*. New York: H. Holt and company (originally published in 1890).
- Janssens, Vera (1999). Over 'slapen' en 'snurken' en de hulp van de context hierbij. *ANBF-nieuwsbrief*, 4: 26-45.
- Jarrold, Christopher, Baddeley, Alan D., Hewes, Alexa K., Leeke, Tansy C. & Phillips, Caroline E. (2004). What links verbal short-term memory performance and vocabulary level? Evidence of changing relationships among individuals with learning disability. *Journal of Memory and Language*, 50(2): 134-148.
- Jiang, Nan & Nekrasova, Tatiana M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3): 433-445.
- Jones, Susan & Sinclair, John (1974). English lexical collocations: A study in computational linguistics. *Cahiers de Lexicologie*, 24(1): 15-61.
- Jonides, J., Schumacher, E. H., Smith, E. E., Koeppel, R. A., Awh, E., Reuter-Lorenz, P. A., Marshuetz, C. & Willis, C. (1998). The role of a parietal cortex in verbal working memory. *Journal of Neuroscience*, 18: 5026-5034.
- Kenny, David A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, 82(6): 887-903.
- Kilgarriff, Adam (2006). BNC database and word frequency lists. Retrieved July 23, 2007, from <http://www.kilgarriff.co.uk/bnc-readme.html>
- Klappenbach, R. (1968). Probleme der phraseologie. *Wissenschaftliche Zeitschrift der Karl-Marx Universität*, 17(5): 221-227.
- Klein-Braley, Christine & Raatz, Ulrich (1984). A cloze-up on the C-test: A study in the construct validation of authentic tests. *Language Testing*, 1: 134-146.
- Kormos, Judith & Sáfár, Anna (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11(2): 261-271.
- Kroll, Judith F. & De Groot, Annette M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In De Groot, A. M. B. & Kroll, J. F. (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah: Lawrence Erlbaum.
- Kroll, Judith F. & Tokowicz, Natasha (2001). The development of conceptual representation for words in a second language. In Nicol, J. L. (Ed.), *One Mind, Two Languages: Bilingual Language Processing*. Oxford: Blackwell.
- Kučera, Henry & Francis, W. Nelson (1967). *Computational Analysis of Present-day American English*. Providence: Brown University Press.
- Kudo, Yoshimitsu (2001). Measuring L2 vocabulary. In Hudson, T. & Brown, J. D. (Eds.), *A Focus on Language Test Development*. Honolulu: University of Hawai'i Press.
- Kurcz, Ida, Lewicki, Andrzej, Sambor, Jadwiga, Szafran, Krzysztof & Woronczak, Jerzy (1990). *Słownik Frekwencyjny Polszczyzny Współczesnej*. Kraków: Instytut Języka Polskiego PAN.
- Lashley, Karl S. (1951). The problem of serial order in behavior. In Jeffress, L. A. (Ed.), *Cerebral Mechanisms in Behavior*. New York: John Wiley.
- Laufer, Batia (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In Schmitt, N. & McCarthy, M. (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- Laufer, Batia (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2): 255-271.
- Laufer, Batia (2005). Lexical Frequency Profiles: From Monte Carlo to the real world - a response to Meara 2005. *Applied Linguistics*, 26(4): 582-588.
- Laufer, Batia & Goldstein, Zahava (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3): 399-436.
- Laufer, Batia & Nation, Paul (1999). A vocabulary-size test of controlled productive ability. *Language Testing* 16(1): 33-51.
- Laufer, Batia & Paribakht, T. Sima (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3): 365-391.
- Leech, Geoffrey (1974). *Semantics*. Harmondsworth: Penguin.

- Leech, Geoffrey, Rayson, Paul & Wilson, Andrew (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman.
- Leśniewska, Justyna & Witalisz, Ewa (2007). Cross-linguistic influence and acceptability judgments of L2 and L1 collocations: A study of advanced Polish learners of English. *EUROSLA Yearbook*, 7(1): 27-48.
- Levy, Betty Ann (1971). Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 10(2): 123-132.
- Lewandowsky, Stephan & Farrell, Simon (2008). Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Language*, 58: 429-448.
- Lewandowsky, Stephan & Murdock, Bennet B. (1989). Memory for serial order. *Psychological Review*, 96(1): 25-57.
- Lewis, Michael (2000). *Teaching Collocations: Further Developments in the Lexical Approach*. Boston: Thomson.
- Lewis, Michael (2002). *The Lexical Approach: the State of ELT and a Way Forward*. Boston: Thomson Heinle.
- Lieven, Elena V.M. & Tomasello, Michael (2008). Children's first language acquisition from a usage-based perspective. In Robinson, P. & Ellis, N. C. (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. London: Routledge.
- Lin, Elizabeth Y. (2009). The Common European Framework and the European Language Portfolio in Taiwan's third-level language education: A longitudinal empirical evaluation. Ph.D. thesis, Trinity College Dublin.
- Little, David (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91(4): 645-655.
- Locke, John (1690). *An Essay Concerning Humane Understanding*. London: Thomas Bassett.
- Louw, Bill (1993). Irony in the text or insincerity in the writer? In Baker, M., Francis, G. & Tognini-Bonelli, E. (Eds.), *Text and Technology*. Amsterdam: Benjamins.
- Lyons, John (1977). *Semantics*. London: Cambridge University Press.
- Mackey, William Francis (1976). *Language Teaching Analysis*. London: Longman.
- MacWhinney, Brian (1999). *The Emergence of Language*. London: Lawrence Erlbaum.
- Martin, Katherine I. (2009). Phonological and working memory and L2 grammar learning. B.A. thesis, University of Michigan.
- Martin, Randi C., Lesch, Mary F. & Bartha, Michael C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, 41: 3-29.
- Masoura, Elvira V. & Gathercole, Susan Elizabeth (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology*, 34(5/6): 383-388.
- Masoura, Elvira V. & Gathercole, Susan Elizabeth (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, 13(3/4): 422-429.
- McAllister, R., Flege, J. E. & Piske, Thorsten (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30: 229-258.
- McCarthy, Michael (1990). *Vocabulary*. Oxford: Oxford University Press.
- McCarthy, Michael & O'Dell, Felicity (2005). *English Collocations in Use*. Cambridge: Cambridge University Press.
- McEnery, Tony, Xiao, Richard & Tono, Yukio (2006). *Corpus-based Language Studies*. London: Routledge Taylor & Francis Group.
- Meara, Paul (1992). *EFL Vocabulary Tests*. University College Swansea: Centre for Applied Language Studies.
- Meara, Paul (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K. & Williams, J. (Eds.), *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press. 35-53.
- Meara, Paul (2005). Lexical Frequency Profiles: A Monte Carlo analysis. *Applied Linguistics*, 26(1): 32-47.
- Meara, Paul & Buxton, Barbara (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2): 142-154.
- Meara, Paul & Fitzpatrick, Tess (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28: 19-30.

- Meara, Paul & Jones, Glyn (1987). Tests of vocabulary size in English as a foreign language. *Polyglot*, 8(1): C1-E14.
- Meara, Paul & Milton, James (2003). *X-Lex, the Swansea Levels Test*. Newbury: Express.
- Mel'cuk, Igor (1998). Collocations and lexical foundations. In Cowie, A. P. (Ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Miller, George A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63: 81-97.
- Milton, James (2006a). French as a foreign language and the Common European Framework of Reference for Languages. *Proceedings from the Crossing Frontiers: Languages and the International Dimension conference* Retrieved from www.llas.ac.uk/cardiff2006.
- Milton, James (2006b). Language lite? Learning French vocabulary in school. *Journal of French Language Studies*, 16: 187-205.
- Milton, James (2006c). X-Lex: The Swansea Vocabulary Levels Test. *Proceedings from the 7th and 8th Current Trends in English Language Testing (CTELT) Conference*, 4: 29-39.
- Milton, James (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Milton, James & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for Languages. In Richards, B., Daller, H. M., Malvern, D., Meara, P., Milton, J. & Treffers-Daller, J. (Eds.), *Vocabulary Studies in First and Second Language Acquisition: The Interface between Theory and Application*. Basingstoke: Palgrave.
- Mochida, Akira & Harrington, Michael (2006). The Yes/No Test as a measure of receptive vocabulary knowledge. *Language Testing*, 32(1): 73-98.
- Montgomery, J. Anne (1977). Sibling intervention in first language acquisition: A case study. Report: ED144384. New Mexico: Dept. of English, New Mexico State University.
- Montgomery, James (1995). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics*, 16: 355-378.
- Moon, Rosamund (1987). The analysis of meaning. In Sinclair, J. (Ed.), *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins ELT.
- Moon, Rosamund (1992). Textual aspects of fixed expressions in learners' dictionaries. In Arnaud, P. & Béjoint, H. (Eds.), *Vocabulary and Applied Linguistics*. London Macmillan.
- Moon, Rosamund (2009a). Sinclair, lexicography, and the Cobuild project: The application of theory. In Moon, R. (Ed.), *Words, Grammar, Text: Revisiting the Work of John Sinclair*. Amsterdam: John Benjamins.
- Moon, Rosamund (2009b). *Words, Grammar, Text: Revisiting the Work of John Sinclair*. Amsterdam: John Benjamins.
- Morris, Lori & Cobb, Tom (2004). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. *System*, 32: 75-87.
- Mueller, Shane T., Seymour, Travis L., Kieras, David E. & Meyer, David E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6): 1353-1380.
- Muncie, James (2002). Process writing and vocabulary development: Comparing Lexical Frequency Profiles across drafts. *System*, 30: 225-235.
- Murdock, Bennet B. (1967). Recent developments in short-term memory. *British Journal of Psychology*, 58: 421-433.
- Murdock, Bennet B. (1974). *Human Memory: Theory and Data*. Hillsdale: Lawrence Erlbaum Associates Inc.
- Murphy, Raymond (2007). *Essential Grammar in Use: Elementary*. Cambridge: Cambridge University Press.
- Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, 21(3): 263-276.
- Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 78(4): 679-684.
- Myles, Florence, Hooper, Janet & Mitchell, Rosamond (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning* 48(3): 323-364.
- Nagy, William, Anderson, Richard C., Schommer, Marlene, Scott, Judith Ann & Stallman, Anne C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24(3): 262-282.
- Nagy, William E. & Anderson, Richard C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19(3): 304-330.

- Nairne, James S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18(3): 251–269.
- Nairne, James S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53(1): 53–81.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5: 12–25.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.
- Nation, I. S. P. (1993). Measuring readiness for simplified material: A test of the first 1,000 words of English. *RELCA Anthology Series*, 31: 193–203.
- Nation, I. S. P. (1998). Should Translations be Tested? Unpublished manuscript.
- Nation, I. S. P. (2001a). How many high frequency words are there in English? In M. Gill, A.W. Johnson, L. M. Kiski, R.D. Sell & Warvik, B. (Eds.), *Language, Learning, Literature: Studies Presented to Hakan Kingdom*. (English Department Publications 4). Turku: Abo Akademi University.
- Nation, I. S. P. (2001b). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In Bogaards, P. & Laufer, B. (Eds.), *Vocabulary in a Second Language*. Amsterdam: John Benjamins Publishing Company.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63(1): 59–82.
- Nation, I. S. P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In Daller, H., Milton, J. & Treffers-Daller, J. (Eds.), *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- Nation, I. S. P. & Beglar, David (2007). A Vocabulary Size Test. *The Language Teacher*, 31(7): 9–13.
- Nation, I. S. P. & Hwang, Kyongho (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1): 35–41.
- Nattinger, James R. & DeCarrico, Jeanette S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nelson, Katherine (1981). Individual differences in language development: Implications for development and language. *Developmental Psychology*, 17(2): 170–187.
- Nesselhauf, Nadia (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Nicolson, R. (1981). The relationship between memory span and processing speed. In Friedman, M., Das, J. P. & O'Connor, N. (Eds.), *Intelligence and Learning*. New York: Plenum Press.
- Nimmo, Lisa M. (2004). Investigating the Phonological Similarity Effect: Implications for Short-term Memory Models. Ph.D. thesis, University of Wollongong. Retrieved, February 6, 2010, from <http://ro.uow.edu.au/theses/457>.
- O'Brien, Irena Socha, Segalowitz, Norman, Collentine, Joe & Freed, Barbara (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, 27: 377–402.
- Osaka, Naoyuki, Logie, Robert H. & D'Esposito, Mark (2007). *The Cognitive Neuroscience of Working Memory*. Oxford: Oxford University Press.
- OUP (2001). *Paper and Pen Test: User Manual*. Oxford: Oxford University Press.
- Oxford Collocations: Dictionary for Students of English (2002). Oxford: Oxford University Press.
- Oxford English Dictionary (2008). Third edition. Oxford: Oxford University Press.
- Page, Michael P. A. & Norris, Dennis (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105(4): 761–781.
- Palmer, Harold E. (1917). *The Scientific Study and Teaching of Languages*. London: G.G. Harrap & Company.
- Palmer, Harold E. (1933). *Second Interim Report on English Collocations*. Tokyo: Kaitakusha (reprinted in 1966).
- Palmer, Harold E. (1934). Director's report. *The Bulletin of the Institute for Research in English Teaching*, 108: 16–24.
- Papagno, Costanza, Valentine, Tim & Baddeley, Alan D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30(3): 331–347.
- Papagno, Costanza & Vallar, Giuseppe (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *The Quarterly Journal of Experimental Psychology*, 44A(1): 47–67.
- Papagno, Costanza & Vallar, Giuseppe (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, 48A(1): 98–107.

- Partington, Alan (1998). *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins Publishing Company.
- Paulesu, E., Frith, C. D. & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362: 342-345.
- Pawley, Andrew & Syder, Frances Hodgetts (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Richards, J. C. & Schmidt, R. W. (Eds.), *Language and Communication*. London: Longman.
- Peters, Ann M. (1977). Language-learning strategies: Does the whole equal the sum of the parts? *Language*, 53(3): 560-573.
- Peters, Ann M. (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Peterson, Lloyd R. & Johnson, Suzanne T. (1971). Some effects of minimizing articulation on short-term retention *Journal of Verbal Learning and Verbal Behavior*, 10(4): 346-354.
- Pickering, Susan J. (2006). Assessment of working memory in children. In Pickering, S. J. (Ed.), *Working Memory and Education*. Oxford: Elsevier.
- Posner, Michael I. & Konick, Andrew F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology*, 72(2): 221-231.
- Pressley, Michael & Van Meter, P. (1992). What is memory development the development of? A 1990s theory of memory and cognitive development 'twix 2 and 20. In Morris, P. & Gruneberg, M. (Eds.), *Theoretical Aspects of Memory*. London: Routledge.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3): 513-536.
- Read, John (1988). Measuring the vocabulary knowledge of second language learners. *RELJ Journal*, 19: 12-25.
- Read, John (1997). Vocabulary and testing. In Schmitt, N. & McCarthy, M. (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- Read, John (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Read, John (2007). Second language vocabulary assessment: Current practices and new directions. *IJES*, 7(2): 105-125.
- Renouf, Antoinette (1987). Corpus development. In Sinclair, J. (Ed.), *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins ELT.
- Revier, Robert Lee (2009). Evaluating a new test of whole English collocations. In Bartfield, A. & Gyllstad, H. (Eds.), *Researching Collocations in Another Language: Multiple Interpretations*. London: Palgrave MacMillan.
- Richards, Brian & Malvern, David (2007). Validity and threats to the validity of vocabulary measurement. In Daller, H. M., Milton, J. & Treffers-Daller, J. (Eds.), *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- Richardson, John T.E. (2007). Measures of short-term memory: A historical overview. *Cortex*, 43: 635-650.
- Rivers, Wilga M. & Temperley, Mary S. (1978). *A Practical Guide to the Teaching of English as a Second or Foreign Language*. Oxford: Oxford University Press
- Romani, Cristina (1994). The role of phonological short-term memory in syntactic parsing: A case study. *Language and Cognitive Processes*, 9(1): 29-67.
- Rost, Michael (1990). *Listening in Language Learning*. London: Longman.
- Salt, J. & Rees, P. (2006). *Globalization, Population Mobility and Impact of Migration on Population*. Swindon: ESRC.
- Schieffelin, Bambi B. & Eisenberg, Ann R. (1981). Cultural variation in children's conversations. In Schiefelbusch, R. L. & Bricker, D. D. (Eds.), *Early Language: Acquisition and Intervention*. Baltimore: University Park Press.
- Schmidt, Richard (1983). Interaction, acculturation and the acquisition of communicative competence: A case study of an adult. In Wolfson, N. & Judd, E. (Eds.), *Sociolinguistics and Language Acquisition: A Book of Readings*. Rowley: Newbury House.
- Schmitt, Norbert (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing*, 16(2): 189-216.
- Schmitt, Norbert (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Schmitt, Norbert (2004). *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins Publishing Company.
- Schmitt, Norbert, Dörnyei, Zoltán, Adolphs, Svenja & Durow, Valerie (2004a). Knowledge and acquisition of formulaic sequences: A longitudinal study. In Schmitt, N. (Ed.), *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins Publishing Company.

- Schmitt, Norbert, Grandage, Sarah & Adolphs, Svenja (2004b). Are corpus-derived recurrent clusters psychologically valid? In Schmitt, N. (Ed.), *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins Publishing Company.
- Schmitt, Norbert & Meara, Paul (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19: 17-36.
- Schmitt, Norbert, Schmitt, Diane & Clapham, Caroline (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1): 55-88.
- Schweickert, Richard (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, 21(2): 168-175.
- Schweickert, Richard, Guentert, Lawrence & Herberger, Lora (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, 1(1): 74-77.
- Seashore, Robert H. & Eckerson, Lois D. (1940). The measurement of individual differences in general English vocabularies. *Journal of Educational Psychology*, 31(1): 14-38.
- Selkirk, E. (1984). On the major class features and syllable theory. In Aronoff, M. & Oehrle, R. T. (Eds.), *Language Sound Structure: Studies in phonology presented to Morris Halle by his teachers and students*. Cambridge: MIT Press.
- Service, Elisabet (1992). Phonology, working memory and foreign-language learning. *The Quarterly Journal of Experimental Psychology*, 45A(1): 21-50.
- Service, Elisabet (2007). Individual differences in phonological learning and verbal STM span. *Memory & Cognition*, 35(5): 1122-1135.
- Service, Elisabet & Craik, Fergus I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, 32(5): 608-623.
- Shallice, Tim & Warrington, Elizabeth K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology*, 22A(2): 261-273.
- Shiffrin, Richard M. & Cook, J. R. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behavior*, 17: 189-218.
- Sinclair, John (1966). Beginning the study of lexis. In Bazell, C. E., Catford, J. C., Halliday, M. A. K. & Robins, R. H. (Eds.), *In Memory of J. R. Firth*. London: Longmans.
- Sinclair, John (1987a). Collocation: A progress report. In Steele, R. & Threadgold, T. (Eds.), *Language Topics: Essays in Honour of Michael Halliday*. John Benjamins Publishing.
- Sinclair, John (1987b). *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins ELT.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John (1996). The search for units of meaning. *Textus*, 1X: 75-106.
- Singleton, David (1999). *Explaining the Second Language Mental Lexicon*. Cambridge: Cambridge University Press.
- Sinn, Hans-Werner (2004). EU Enlargement, Migration and the New Constitution. CESifo Working Paper No. 1367.
- Siyanova, Anna & Schmitt, Norbert (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes*, 64(3): 429-458.
- Smith, Bruce (2006). Precautions regarding nonword repetition tasks. *Applied Psycholinguistics*, 27(4): 584-587.
- Smyth, Mary M., Pearson, Norma A. & Pendleton, Lindsay R. (1988). Movement and working memory: Patterns and positions in space. *The Quarterly Journal of Experimental Psychology*, 40A(3): 497 - 514.
- Snowling, Maggie, Chiat, Shula & Hulme, Charles (1991). Words, nonwords and phonological processes: Some comments on Gathercole, Willis, Emslie and Baddeley *Applied Psycholinguistics*, 12(3): 369-373.
- Soars, Liz & Soars, John (2006). *New Headway: Elementary*. Oxford: Oxford University Press.
- Speciale, Giovanna, Ellis, Nick C. & Bywater, Tracey (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, 25(2): 293-321.
- Speidel, Gisela E. (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study *First Language*, 13(37): 69-91.
- Takala, S. (1984). *Evaluation of Students' Knowledge of English Vocabulary in the Finnish Comprehensive School*. Jyväskylä: Jyväskylän Yliopisto.
- Tharp, James B. (1934). The basic French vocabulary and its use. *The Modern Language Journal*, 19(2): 123-131.

- Thorn, Annabel S. C., Gathercole, Susan Elizabeth & Frankish, Clive R. (2002). Language familiarity effects in short-term memory: The role of output delay and long-term knowledge. *The Quarterly Journal of Experimental Psychology*, 55A: 1363-1383.
- Thorndike, Edward L. & Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University.
- Tomasello, Michael (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. London: Harvard University Press.
- Trojano, Luigi, Stanzione, Maria & Grossi, Dario (1992). Short-term memory and verbal learning with auditory phonological coding defect: A neuropsychological case study. *Brain and Cognition*, 18(1): 12-33.
- Tschirner, Erwin (2004). Breadth of vocabulary and advanced English study: An empirical investigation. *Electronic Journal of Foreign Language Teaching*, 1(2): 27-39.
- Tulving, Endel & Craik, Fergus I. M. (2005). *The Oxford Handbook of Memory*. Oxford: Oxford University Press.
- Underwood, Geoffrey, Schmitt, Norbert & Galpin, Adam (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In Schmitt, N. (Ed.), *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins Publishing Company.
- Vallar, Giuseppe & Baddeley, Alan D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, 23: 151-161.
- Vallar, Giuseppe & Papagno, Costanza (2002). Neuropsychological impairments of verbal short-term memory. In Baddeley, A. D., Kopelman, M. D. & Wilson, B. A. (Eds.), *Handbook of Memory Disorders*. Chichester: Wiley.
- Vallar, Giuseppe & Shallice, Tim (1990). *Neuropsychological Impairments of Short-term Memory*. Cambridge: Cambridge University Press.
- van Ek, Jan Ate & Trim, J. L. M. (1990a). *Threshold 1990: Council Of Europe*. Strasbourg: Council of Europe Publishing.
- van Ek, Jan Ate & Trim, J. L. M. (1990b). *Waystage 1990: Council Of Europe*. Strasbourg: Council of Europe Publishing.
- Vermeer, Anne (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In Bogaards, P. & Laufer, B. (Eds.), *Vocabulary in a Second Language*. Amsterdam: John Benjamins Publishing Company.
- Verstraten, Linda (1992). Fixed phrases in monolingual learners' dictionaries. In Arnaud, P. J. L. & Béjoint, H. (Eds.), *Vocabulary and Applied Linguistics*. London: Macmillan.
- Vierra, Andrea & Pollock, Judith (1992). *Reading Educational Research*. Scottsdale: Gorsuch Scarisbrick.
- Vitevitch, Michael S., Luce, Paul A., Charles-Luce, Jan & Kemmerer, David (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40(1): 47-62.
- Waring, Rob (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata: The Occasional Papers at Notre Dame Seishin University*: 94-114.
- Waters, Gloria S. & Caplan, David (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103(4): 761-772.
- Waugh, Nancy C. & Norman, Donald A. (1965). Primary memory. *Psychological Review*, 72(2): 89-104.
- Webb, Stuart (2008). Receptive and productive vocabulary sizes of L2 Learners. *Studies in Second Language Acquisition*, 30: 79-95.
- Weinert, Regina (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16(2): 180-205.
- Weinreich, Uriel (1953). *Languages in Contact*. New York: Linguistic Circle of New York.
- Weinreich, Uriel (1969). Problems in the analysis of idioms. In Puhvel, J. (Ed.), *Substance and Structure of Language*. Berkeley: University of California Press.
- Wesche, Marjorie & Paribakht, T. Sima (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53: 13-40.
- West, Michael (1953). *A General Service List of English Words*. London: Longman.
- Williams, John N. & Lovatt, Peter (2005). Phonological memory and rule learning. *Language Learning*, 55(1): 177-233.
- Woutersen, Mirjam (1997). *Bilingual Word Perception*. Nijmegen: Katholieke Universiteit Nijmegen.
- Wray, Alison (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4): 463-489.

- Wray, Alison (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Xiao, Richard & McEnery, Tony (2006). Collocation, semantic prosody, and near-synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1): 103-129.
- Xing, Peiling (2007). Lexis learning rainbow: A longitudinal study of Chinese learners' vocabulary acquisition strategies in UK universities. Ph.D. thesis, University of Dundee.
- Xing, Peiling & Fulcher, Glenn (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, 35: 182-191.
- Xue, G. & Nation, I.S.P. (1984). A University Word List. *Language Learning and Communication*, 3: 215-229.