An exploration of second language collocation knowledge and development.

Barfield, Andrew William

An Exploration of Second Language Collocation Knowledge and Development

Andrew William Barfield

Submitted to the University of Wales in fulfillment of the requirements for the Degree of Doctorate of Philosophy

University of Wales Swansea

2006

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Despite the wealth of L1 collocation studies and the explosive growth in corpus linguistics in the last 20 years, relatively little research has been conducted into L2 collocation knowledge and development. Most L2 studies have involved isolated single interventions that have taken as their main research focus collocation errors or collocation types. The assumption is that (advanced) L2 learners should strive for native-like collocation ability and that their lack of collocation accuracy should be judged by a far-removed NS standard. Extremely little experimental work has been completed on the development of L2 collocation knowledge at lower levels of proficiency. Even less has been carried out into how learners themselves address developing their L2 collocation knowledge and what psycholinguistic and contextual factors might be involved. This research sets out to investigate how we can experimentally measure L2 collocation recognition knowledge and production knowledge. It also seeks to examine factors in learning that help or hinder the development of L2 collocation knowledge. From exploring L2 collocation knowledge from these three viewpoints, an experimentally grounded model of L2 collocation knowledge is proposed. This model has several implications for how we might understand the development and organization of the L2 lexicon in relation to L2 collocation knowledge.

Declarations and Statements

DECLARATION

This work has not previously been submitted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... **(candidate)**

Date ..... September 20th 2006

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date ..... September 20th 2006

STATEMENT 2

I hereby give my consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available to outside organizations.

Signed ..... (candidate)

Date ..... September 20th 2006

Acknowledgements

My deepest thanks go to Professor Paul Meara, my supervisor, for all his support, criticism, good humour and academic rigour. Paul has been intellectually demanding and lexically stimulating, and I feel extremely fortunate to have been his student. I would like to thank Paul for his truly inspirational teaching and research. Most importantly, as a supervisor, he was always there, which I think is quite remarkable. Thank you, Paul, for everything you have done to help me.

Out of many, one colleague and friend deserves special thanks, John Shillaw. Over the years, John has been a constant source of support and wise counsel on matters lexical as much as life in general. I'd like to acknowledge my deep sense of gratitude and friendship to John.

Doing a distance PhD at Swansea also means appreciating the contribution of many different CALS staff members. I would simply like to say a huge thank you to everybody working in CALS for being such a motivated/-ing group of teachers and researchers. I would also like to voice my thanks to my fellow PhD students and 'senpai' who have shared fascinating questions, feedback and help. In particular, I would like to thank Andy Mellor, Andy Wilcox, Brent Wolter, David Coulson, Jim Ronald, Masa Mochizuki, Patrick McGavigan, Richard Pemberton, Simon Fraser, Rob Waring, and Tad Kamimoto.

I am also very happy to record my thanks to friends and colleagues who have at different times helped the conversations about collocation continue. Here, I would like to mention Akiko Kawasaki, Alan Milne, Henrik Gyllstad, Hiro Iwasaki, Hugh Nicoll, Mamoru Kasahara, Marijana Perišić, Meg Arai, Mike Nix, Neil Cowie, Ramesh Krishnamurty, Richard Smith, Richard Walton, Shizuo Satoma, Steve Brown, Steve Hesse, Takeshi Yuhara, Tim Ashwell, Tim Knowles, Yuji Kaneko, Yukio Saigusa, Yuri Komuro and Zorana Vasiljević. My Chuo colleagues have been extremely generous of their time at work so that I could find space for my research: thank you to all of them for their goodwill and support.

My thanks also go to the many students and colleagues who took part in the different studies that I did. In addition, I'd like to recognize the great assistance provided by my TAs, Norie Kumakura and Hiko Jyo.

On the personal side, let me thank my family for their quiet encouragement: my parents, my brother and his family in England; Tin Tin's parents, her brother and sister in Burma; and my daughter Ana (in Wales, Turkey, Serbia, America and China!).

Last but by no means least, I say the most personal thank you to my partner in life, Tin Tin Htun, and our son, Moe Min Michael. Neither words nor collocations could ever capture what I feel in recognition of your loving kindness, so I simply dedicate this thesis to both of you.

List of Contents 5
List of Appendices 10
List of Tables 11
List of Figures 14

Chapter 1 Two's Company, Three's A Crowd? Interpretations and Issues in Researching Collocation 15
1.0 Overview 15
1.1 Making sense of collocation: evolution of the term and concept 15
1.2 The phraseological interpretation of collocation 18
1.2.1 Palmer and the phraseological approach to collocation 18
1.2.2 Building on the Palmer legacy: Hornby 22
1.2.3 Extending the phraseological interpretation: Cowie 24
1.2.4 Summing up the phraseological interpretation 26
1.3 The typological interpretation of collocation 27
1.3.1 The Russian tradition 27
1.3.2 Further restrictions on lexical combinations 29
1.3.3 The semantic labeling of lexical collocations and lexical functions 32
1.3.4 Summing up the typological interpretation 34
1.4 The textual interpretation of collocation 35
1.4.1 Firth and collocation as a level of language 35
1.4.2 Firth's theory applied: Halliday's interpretation of collocation 37
1.4.3 Firth's theory applied: Sinclair's interpretation of collocation 39
1.4.4 Summing up the textual interpretation 45
1.5 Drawing things together: Two *is* company and three *is* a crowd, but 46

Chapter 2 Literature Review of Second Language Collocation Research 48
2.0 Overview 48
2.1 Phrase- & sentence-level analysis 49
2.1.1 Study 1: Biskup (1992) 49
2.1.2 Study 2: Bahns & Eldaw (1993) 52
2.1.3 Study 3: Farghal & Obiedat (1995) 55
2.1.4 Study 4: Bonk (2000) 58
2.1.5 Concluding comments on the phrase- & sentence-level analysis studies. 61
2.2 Prompted essay with large-scale collocation inventory analysis 62
2.2.1 Study 5: Zhang (1993) 62
2.2.2 Study 6: Gitsaki (1996) 65
2.2.3 Concluding comments on the large-scale collocation inventory analysis studies 68
2.3 Essay corpus with narrow analysis of specific collocation types 68
2.3.1 Study 7: Dechert & Lennon (1989) 68
2.3.2 Study 8: Chi, Wong & Wong (1994) 72
2.3.3 Study 9: Granger (1998) 75
2.3.4 Study 10: Howarth (1998) 78
2.3.5 Study 11: Nesselhauf (2003) 81

2.3.6	Concluding comments on the essay corpus with narrow analysis of specific collocation types studies	84
2.4	Experimental measures of collocation knowledge	85
2.4.1	Studies 12 and 13: Read (1993, 1998)	85
2.4.2	Study 14: Schmitt (1999)	89
2.4.3	Study 15: Mochizuki (2002)	92
2.4.4	Concluding comments on the experimental measures of collocation knowledge studies	95
2.5	Discussion	98
2.5.1	The focus on collocation types	98
2.5.2	Collocation production and recognition	99
2.5.3	Collocation knowledge and other types of lexical knowledge and proficiency	100
2.5.4	Learning collocations	101
2.6	Conclusion	101
Chapter 3	**Exploring the Collocation Environments of Lexical Verbs**	**103**
3.0	Overview	103
3.1	Introduction	103
3.2	Method	106
3.3	Results	109
3.3.1	Comparison of groups	109
3.3.2	Lexical verbs common to all essays	109
3.3.3	A small set of high frequency lexical verbs	112
3.3.4	A small set of low frequency lexical verbs	112
3.4	Discussion	115
3.5	Conclusion	121
Chapter 4	**A Preliminary Exploration of Lexical Verb Knowledge**	**123**
4.0	Overview	123
4.1	Introduction	123
4.2	The Vocabulary Levels Test	124
4.3	Research goals	127
4.4	Research construct and design	128
4.5	Method	130
4.6	Descriptive statistics	131
4.7	Results	132
4.7.1	Item-total correlations	132
4.7.2	Differences in results for a high group and a low group	134
4.7.3	Correlations with overall English proficiency	137
4.7.4	Insights from the think-aloud interviews	137
4.8	Discussion	141
4.8.1	Differences between most-known and least-known verbs	141
4.8.2	The problem of construct validity	143
4.8.3	Item independence and learners' lexical knowledge	145
4.9	Conclusion	146

Chapter 5	**Measuring States of Lexical Verb Knowledge**	**149**
5.0	Overview	149
5.1	Introduction	150
5.2	Interpreting complex partial knowledge decision-making	152
5.2.1	Complex partial knowledge decision-making	152
5.2.2	The vocabulary scale	153
5.2.3	Weaknesses of the vocabulary scale	156
5.2.4	An alternative to the vocabulary scale	158
5.3	Research goals	160
5.4	Research method	160
5.5	Results	163
5.5.1	Results: descriptive statistics	163
5.5.2	Results: comparison of subjects for the four knowledge states	164
5.5.3	Results: comparison of different groups	164
5.5.4	Results: rank order of lexical verbs	166
5.6	Discussion	170
5.6.1	Evaluating the test instrument	170
5.6.2	Learner recognition of lexical verbs	171
5.6.3	The relationship between learner recognition and verb frequency	173
5.6.4	Drawing the discussion together	176
5.7	Conclusion	176

Chapter 6	**Exploring Noun and Verb + Noun Recognition**	**179**
6.0	Overview	179
6.1	Introduction	180
6.2	Research questions	181
6.3	Research design and method	181
6.3.1	Design of the VN collocation test	185
6.3.2	Design of the noun recognition test	188
6.3.3	Research method	188
6.4	Results for the noun recognition test	188
6.4.1	Noun recognition: descriptive statistics	189
6.4.2	Noun recognition: word-based analysis results	190
6.4.3	Noun recognition: subject-based analysis results	190
6.4.4	Drawing together the noun recognition results	191
6.5	Method for the VN recognition test	191
6.6	Results for the VN recognition test	192
6.6.1	Real collocations: descriptive statistics	193
6.6.2	Real collocations: rank order results	193
6.6.3	Real collocations: chi-square analysis results	196
6.6.4	Real collocations: collocation recognition and verb/noun knowledge results	197
6.6.5	Real collocations: comparison of groups results	200
6.6.6	Drawing together the real collocation recognition results	200
6.7	Results: mis-collocations	201

6.8	Real collocations: correlations between verb, noun, and collocation totals for Knowledge State 4	203
6.9	Discussion	204
6.10	Conclusion	208
Chapter 7	**Investigating Individual Variation in Collocation Production over Time**	**210**
7.0	Overview	210
7.1	Introduction	210
7.2	Areas of focus	213
7.3	Areas of exploration	215
7.4	The research instrument	216
7.5	Subjects, procedure and scoring	216
7.6	Word association results	220
7.6.1	Lexical profiling of the word associations	220
7.6.2	Total word associations	221
7.6.3	Syntagmatic word associations	224
7.7	Collocation results	227
7.7.1	Collocation profiles	228
7.7.2	Total different appropriate collocations	229
7.7.3	Repetition of different appropriate collocations	230
7.7.4	Mis-collocations	232
7.7.5	Overlaps in appropriate collocations	233
7.7.6	Appropriate collocate types and frequencies	233
7.8	Discussion	236
7.9	Limitations	240
7.10	Conclusion	241
Chapter 8	**Developing a Productive Test of Collocation Knowledge**	**243**
8.0	Overview	243
8.1	Introduction	243
8.2	Research focus	244
8.3	Method	244
8.3.1	Selection and norming of the target items	245
8.3.2	Test procedure and scoring	247
8.3.3	Analytical approach	249
8.4	Results	253
8.4.1	Descriptive statistics for collocation appropriacy	253
8.4.2	High and low group lexical profiles for all collocates	254
8.4.3	High and low group lexical profiles for appropriate collocates only	255
8.4.4	Differences between high and low group collocate responses	255
8.4.5	The relationship between person ability and item difficulty	257
8.4.6	Summing up the results	263
8.5	Discussion	263
8.6	Conclusion	267

Chapter 9 Exploring Individual Development of Collocation Knowledge..... 269
9.0 Overview..... 269
9.1 Introduction..... 270
9.2 Research focus..... 275
9.3 Method..... 275
9.3.1 Participants..... 275
9.3.2 Instrument: organisation of interviews..... 277
9.3.3 Overview of grounded theory analysis..... 277
9.3.4 Stages of analysis..... 279
9.4 Results..... 282
9.4.1 Care..... 283
9.4.2 Engagement..... 285
9.4.3 Reading..... 286
9.4.4 Vocabulary selection..... 287
9.4.5 Vocabulary connections..... 289
9.4.6 Dictionary use..... 294
9.4.7 Noticing..... 296
9.4.8 Problematization of learning..... 297
9.4.9 Emerging theorization of collocation development..... 300
9.5 Discussion..... 301
9.6 Conclusion..... 304

Chapter 10 Discussion..... 306
10.0 Overview..... 306
10.1 Gaps in exploring L2 collocation knowledge..... 306
10.2 Summary of insights..... 307
10.2.1 Insights concerning collocation recognition..... 310
10.2.2 Insights concerning collocation production..... 312
10.2.3 Insights concerning individual collocation process..... 315
10.3 Issues for discussion..... 316
10.4 Exploring dimensions of collocation knowledge and use..... 317
10.4.1 The tension between delexicalisation and lexicality in L2 collocation use..... 317
10.4.2 The combination of frequent and infrequent lexical items in L2 collocation use..... 324
10.4.3 The centrality of noun knowledge in L2 collocation knowledge and implications for L2 collocation development..... 329
10.4.4 Summing up the discussion on exploring dimensions of collocation knowledge and use..... 334
10.5 Linking lexical knowledge and use with collocation development..... 334
10.5.1 The bilingual lexicon: network development and conceptual shift..... 335
10.5.2 Differences between native-like and L2 collocation ability..... 341
10.5.3 The priming role of associative knowledge and paraphrase for L2 collocation development..... 346
10.5.4 Summing up linking lexical knowledge and use with collocation development..... 349

10.6 Proposing a model of L2 collocation development..... 351
10.6.1 An emerging model of L2 collocation development..... 352
10.7 Conclusion..... 354

Chapter 11 Conclusion..... 356

References..... 362

Appendices..... CD-ROM
1.1 An example lexical constellation (Palmer, 1931, p.33)..... 375
1.2 An example associative word family (Hornby, 1947, p.94)..... 375
1.3 A sample lexical entry, ECD style (Mel'ćuk, 1998, pp.52-53)..... 376
1.4 A network of collocations for expressing ideas of time (Jones & Sinclair, 1974, p.41)..... 377
2.1 37 types of collocation investigated by Gitsaki (1996)..... 378
2.2 The 5-level classification of collocation restrictions proposed by Howarth (1998)..... 379
2.3 The 3-way classification of collocation restrictions proposed by Nesselhauf (2003)..... 380
2.4 Modified Vocabulary Knowledge Scale used in Read (1998)..... 381
2.5 Example NS collocation responses to two prompts in Schmitt (1999)..... 382
3.1 Occurrence of lexical verbs across 38 essays..... 383
4.1 The University Word List Verb Test..... 384
6.1 Chi-square analysis results..... 387
7.1 Example association/collocation networks for T1-T6 for two items..... 395
8.1 Native speaker & non-native speaker verb collocates for 50 nouns..... 398
8.2 Frequent and infrequent collocates for 30 nouns..... 403
8.3 Collocate database for 30 nouns..... 404
9.1 Interview transcripts..... 416
9.2 Visual displays of certain other key categories..... 463
9.2A Engagement with content in Kensuke's account
9.2B Engagement with content in Shoichi's account
9.2C Engagement with content in Emi's account
9.2D Noticing in Emi's account

List of Tables

Table 1.1	Differential codification of example verb + object noun collocations (based on Cowie, 1999, p.60)	23
Table 1.2	Howarth's criteria for classifying V+N combinations	31
Table 1.3	Howarth's semantic division of verb types	31
Table 1.4	A model for reconciling the syntagmatic and paradigmatic axes	43
Table 1.5	The syntagmatic-paradigmatic model applied to **budge**	44
Table 2.1	Classification of previous L2 collocation studies	48
Table 2.2	Target items (Bonk, 2000)	60
Table 2.3	NNS lexical collocations results in Howarth (1998)	79
Table 2.4	Key-feature summary of previous research into L2 collocation knowledge	97
Table 3.1	T-unit results for the higher proficiency group	108
Table 3.2	T-unit results for the lower proficiency group	108
Table 3.3	Concordances for **want***	110
Table 3.4	Most common collocates of **want** occurring one word to the left and to the right	111
Table 3.5	Most common collocates of **want to**	111
Table 3.6	Concordances for four high frequency lexical verbs	112
Table 3.7	Occurrences of UWL verbs	113
Table 3.8	Immediate environment collocates for **research** (closest noun phrase to the left and right)	113
Table 3.9	Immediate environment noun phrase collocates for UWL verbs (with UWL items highlighted in bold)	114
Table 3.10	Structural classification of UWL verb noun phrase collocates	118
Table 4.1	Item-total correlation in Beglar and Hunt's validation	126
Table 4.2	Descriptive statistics: UWL verb-test compared to Beglar and Hunt revised UWL forms	131
Table 4.3	Overview of item-total correlations in the UWL-verb test	133
Table 4.4	Individual item-total correlations in the UWL-verb test	134
Table 4.5	Comparison of chi-square results for the low and high groups	135
Table 4.6	Comparison of most-known and least-known verbs	136
Table 4.7	Correlations between academic verb knowledge and general English proficiency	137
Table 4.8	Use of absolute knowledge vs. partial knowledge in completing the UWL verb test	138
Table 4.9	Absolute vs. test-taking and lexical strategies in completing the UWL verb test	139
Table 4.10	Specific lexical strategies in completing the UWL verb test	139
Table 4.11	Each interviewee's use of specific lexical strategies	140
Table 4.12	Word-length: most-known and least-known verbs	141
Table 4.13	BNC frequencies and UWL frequencies & ranges for the most-known and least-known verbs	143
Table 5.1	Discrete types of word knowledge (Nation, 2001, p.347)	151
Table 5.2	The problem of interpreting VKS scores	157

Table 5.3	Verbs selected for knowledge-state ranking	162
Table 5.4	Lexical verbs assigned to each knowledge state	163
Table 5.5	Comparison of total scores for the four knowledge states (*p* values)	164
Table 5.6	Verbs in rank order for Knowledge State 4, with source frequency range and BNC rank (N=217)	166
Table 5.7	12 strongly recognized verbs outside the BNC 2000 most frequent range	172
Table 5.8	11 weakly recognized verbs inside the BNC 2000 most frequent range	172
Table 5.9	Cumulative distributions of source ranges	174
Table 5.10	BNC 2000 and GSL 2000 coverage of the 50 most known verbs	175
Table 6.1	40 verbs selected for the VN collocation test	184
Table 6.2	20 mis-collocations selected for the VN collocation test	185
Table 6.3	120 VN collocations selected (with mis-collocations shown in bold italics)	186
Table 6.4	Nouns assigned to different knowledge states	189
Table 6.5	Number of subjects rating nouns in Knowledge State 4 and Knowledge States 3 & 4 combined, with total <74 (N=93)	190
Table 6.6	Descriptive statistics for real collocation recognition	193
Table 6.7	Rank order results for real collocations for Knowledge State 4 and Knowledge States 3 & 4 combined (N=93)	194
Table 6.8	Example chi-square results: **acquire reputation**	196
Table 6.9	Collocations showing significance of association	197
Table 6.10	Vn, Nv, and vn is greater than VN for recognised collocations	198
Table 6.11	Number of subjects knowing verb, noun and collocation	199
Table 6.12	Mis-collocations assigned to each knowledge state	201
Table 6.13	Mis-collocations claimed as known	202
Table 6.14	Lexical profile of the 15 most recognized collocations	205
Table 6.15	Lexical congruence of the 15 most recognized collocations	206
Table 7.1	Example word associations and collocations for **advice**	215
Table 7.2	Database of appropriate collocates for **advice**	219
Table 7.3	Infrequent lexical profiles of word associations	220
Table 7.4	Total different word associations	222
Table 7.5	Total number of different associations produced T1-T6	223
Table 7.6	Emi's syntagmatic word association responses	225
Table 7.7	Kensuke's syntagmatic word association responses	226
Table 7.8	Shoichi's syntagmatic word association responses	227
Table 7.9	Appropriate collocation profiles	228
Table 7.10	Total different appropriate collocations	230
Table 7.11	Distribution of different appropriate collocations	231
Table 7.12	Mis-collocation totals	232
Table 7.13	Overlaps in appropriate collocations	233
Table 7.14	Acceptable collocates produced by Emi T1-T6	234
Table 7.15	Acceptable collocates produced by Kensuke T1-T6	234
Table 7.16	Acceptable collocates produced by Shoichi T1-T6	235
Table 7.17	Lexical profile of acceptable collocates	235
Table 7.18	Range of infrequent acceptable collocates	236

Table 7.19	Grammatical profile of all acceptable collocates	236
Table 8.1	50 nouns selected for piloting	245
Table 8.2	Example non-divergent and divergent items	246
Table 8.3	30 nouns selected from NS & NNS piloting	246
Table 8.4	Native speaker norming bio-data summary	247
Table 8.5	Example responses and scoring	249
Table 8.6	Descriptive statistics	253
Table 8.7	Frequent collocate response totals for both groups	254
Table 8.8	Low and high group appropriate collocate scores	255
Table 8.9	Lexical profile of frequent appropriate collocate responses	255
Table 8.10	Lexical profile of infrequent appropriate collocate responses	255
Table 8.11	High and low group means: all appropriate collocates	256
Table 8.12	High and low group means: frequent appropriate collocates	256
Table 8.13	Differences between high and low means for word class choice	256
Table 8.14	Item fit results for COLLPROD	258
Table 8.15	Person fit results for COLLPROD	259
Table 8.16	Person misfit results for COLLPROD (infit mean square < .65)	261
Table 9.1	Grouping and definition of core categories	281
Table 10.1	Summary of insights from each experimental chapter	308
Table 10.2	Issues for further discussion	316
Table 10.3	Delexicalization and lexicality: a hypothetical example	320
Table 10.4	Conventionalized collocation knowledge and vocabulary size: a hypothetical example	321
Table 10.5	Learner collocation knowledge and vocabulary size: a hypothetical example	321
Table 10.6	Top 20 collocates of **address** (in order of frequency)	327
Table 11.1	Example items for a future adaptation of COLLPROD	358

List of Figures

Figure 5.1	VKS scoring categories (Paribakht & Wesche, 1996, p.30)	155
Figure 5.2	Four states of lexical knowledge (based on Meara, 1999)	159
Figure 5.3	Group mean scores for Knowledge State 4	165
Figure 5.4	Group mean scores for Knowledge States 3 & 4 combined	165
Figure 6.1	Scatterplot for the correlation between noun and verb knowledge	203
Figure 7.1	Lexical profiles of word associations over time	221
Figure 7.2	Collocation profiles over time	229
Figure 7.3	Comparison of collocation profiles T1-T3 and T4-T6	238
Figure 8.1	Item-ability map for COLLPROD	262
Figure 9.1	Care for overall process in Kensuke's account	283
Figure 9.2	Care for overall process in Shoichi's account	284
Figure 9.3	Care for overall process in Emi's account	284
Figure 9.4	Vocabulary selection in Kensuke's account	287
Figure 9.5	Vocabulary selection in Shoichi's account	288
Figure 9.6	Vocabulary selection in Emi's account	289
Figure 9.7	Vocabulary connections in Kensuke's account	290
Figure 9.8	Vocabulary connections in Shoichi's account	291
Figure 9.9	Vocabulary connections in Emi's account	293
Figure 9.10	Tensions emerging in a theorisation of L2 collocation knowledge development	300
Figure 10.1	L1 conceptual mediation: L1-linked collocation	338
Figure 10.2	Mixed L1/L2 conceptual mediation: mixed L1/L2 collocation	339
Figure 10.3	Conceptual knowledge mediated by the L2: L2-linked collocation	339
Figure 10.4	An emerging model of variable L2 collocation development	353
Figure 11.1	Example display of L2 items for linking	359
Figure 11.2	Example display of possible collocation links	359

Chapter 1
Two's Company, Three's A Crowd?
Interpretations and Issues in Researching Collocation

1.0 Overview

Just over 70 years ago, the term **collocation** had been barely used in language studies—and where it had, it was within the realms of descriptive grammar. Yet, since then, collocation has come to take up a central position in how we understand lexis within language use. Most of our contemporary knowledge has come from theorizing and investigating what L1 users do with collocation; far less research has been conducted into what second language learners know about collocations and how they approach the business of learning collocations in another language. Although we are concerned in this thesis with understanding the development of second language collocation knowledge, we start by broadly examining collocation studies and outlining distinct interpretations of collocation. This will allow us to raise issues about researching collocation, as well as position the investigation of the acquisition of second language collocation knowledge against that broader background.

1.1 Making sense of collocation: evolution of the term and concept

The term **collocation** is now defined in modern English dictionaries with a striking degree of overlap. The Longman Dictionary of Contemporary English offers the following definition: "the way in which some words are often used together, or a particular combination of words used in this way: "Commit a crime" is a typical collocation in English" (Longman, 1995, p.258). The Oxford Advanced Learner's Dictionary defines collocation as "a combination of words in a language that happens very often and more frequently than would happen by chance: 'Resounding success' and

'crying shame' are English collocations" (Oxford University Press, 2000, pp.233-234), while collocation is "...the way that some words occur regularly whenever another word is used...the basic notion of collocation" according to the Collin's COBUILD Advanced Learner's English Dictionary (Collins, 2006, p.265). The current consensus nevertheless obscures how the term collocation was historically used in different ways from the 17th century onwards—and, more importantly, how quite distinct theoretical interpretations of the concept have been produced in the last 70 years.

Recorded by Johnson (1755) in an example usage by Bacon in his *Natural History* in 1627 ("In the collocation of the spirits and bodies, the collocation is equal or unequal; and the spirits are coacervate or diffused"), the term collocation seems to have been first accorded a linguistic sense by Harris in 1750: "The accusative—in modern languages...being subsequent to its verb, in the collocation of the words" (Palmer, 1933b, p.7). In the late nineteenth century, Sweet applied the word in a more familiar modern usage: "From such a collocation as **the round earth** we can infer the statement **the earth is round**" (Sweet, 1891, p.17). However, collocation was neither highlighted nor further elaborated by Sweet as a category in linguistic classification.

Such elaboration came 40 years later when Palmer selected collocation[1] as a specialized linguistic term for "units of words that are more than single words" (1931, p.4) and for "any coming-together of words whether to be learnt as integral wholes or susceptible of free compoundings" (1933b, p.7). Whereas Palmer stressed the phraseological character of collocations—"successions of words that must or should be learnt, or are best learnt as integral wholes rather than pieced together from their component parts" (Palmer, 1933b, p.10)—others emphasized the combinatory processes of collocation. Jespersen, for example, used the term collocation for describing how

primary and secondary substantives are combined in phrases such as "**garden flower**" and "**flower garden**" (Jespersen, 1933, p.376). This was primarily a grammatical view of word class combination that was later developed into a concern with collocation categories and types. In the 1940s, Trager, for example, took collocation to relate the form of Russian noun declensions to function and to establish animate and inanimate noun classes: "Collocation establishes categories by stating the elements with which the element being studied enters into possible combinations" (Trager, 1940, p.301). Research was also conducted in Russia into how different types of verbs and nouns are combined and how different types of combinations are semantically motivated (e.g., Vinogradov, 1947; Asomova, 1963). This second approach can be characterized as the typological interpretation of collocation. The typological approach was most obviously applied in later years by lexicographers working on dictionaries of English word combinations (e.g., Benson, Benson & Ilson, 1986b) and theoreticians (e.g., Mel'čuk, 1998) developing explanatory and combinatory dictionaries. A third approach to collocation can be seen originating in the work of Firth in the 1950s. Firth focused on habitual lexical associations in text and saw collocation as a level of language use within a theory of language description. The Firthian interpretation of collocation as part of the system of meaning production was in due course taken up by Halliday and Sinclair in the 1960s and extensively expanded. It has since come to characterize what is arguably the mainstream lexical, text-based view of collocation in contemporary applied English language studies.

This opening sketch of the evolution of the term and concept points us, then, towards three major interpretations of collocation—the phraseological, the typological, and the textual—in applied linguistics. To gain a clearer sense of these three approaches, we will continue by looking at each respective interpretation in more detail.

1.2 The phraseological interpretation of collocation

As noted above, it was Palmer who coined the term collocation in an initial applied linguistic sense, but until recently he has been rarely acknowledged, if at all, for his innovative work on collocation. Firth makes no mention of Palmer's contribution, while Mitchell footnotes a regretfully incomplete reference:

> The term was originally Firth's and he may well have been influenced in the selection by H.E. Palmer, who from Tokyo wrote a monograph on the subject in the nineteen-thirties (a work which I have seen but which I am at present unable to locate and whose title unhappily escapes me).
> (Mitchell, 1971, p.35)

The record has since become a little more precise. Krishnamurty, for example, observes: "The idea of collocation first emerged in the work of language teachers between the two world wars, particularly that of Harold Palmer in Japan" (Krishnamurty, 2004, p.ix; see also Nation, 2001, p.317). Given this generally limited presentation of Palmer's contribution, it is worth looking first at his ideas about collocation more closely. Here, we will also consider the related work of Hornby and Cowie.

1.2.1 Palmer and the phraseological approach to collocation

Palmer's work was driven by a concern with vocabulary selection and control so that appropriate limited vocabularies could be created for producing simplified texts for English learners in Japan (1933b, p.11). Although his study of collocations had this longer-term aim, Palmer saw the *Second Interim Report on Collocations* (Palmer, 1933b) primarily as a technical tool for informing pedagogic decisions about English vocabulary: "...it is essential that those who cater for the needs of foreign students should have at their disposal corresponding lists of collocations" (Palmer, 1931, p.21). His work on collocations was directly related to the 1931 *Second Interim Report on*

Vocabulary Selection, which had put forward a general word list of 3,000 individual items. Eschewing purely objective word counts, Palmer argued that individual vocabulary items that had "common collocations" were "to be given a higher credit than words that do not so occur" (Palmer, 1931, p.16). The general word list was, in its final stages, divided into five sets of 600 words which could correspond to "a five-year secondary school course" (Palmer, 1931, p.33), but which were "incomplete without a corresponding list of 'collocations'" (Palmer, 1931, pp.4-5).

The 1933 report answered this need by collecting together and classifying English collocations that occurred with individual items in the general word list. Collocations were grouped under the main headings of verb collocations, noun collocations, determinative collocations, adjective collocations, adverb collocations, preposition collocations, and connective collocations. These word class-based categories were sub-divided into more specific groupings. For example, one of the verb collocation sub-categories covered "all combinations of verbs with specific nouns." This sub-category started with to **acknowledge receipt + of + N_3**, included **to do justice + to + N_3** and **to take interest [no, little, *etc.* interest] (+ in + N_3)**, and ended with **to wish [+ N_3 +] good luck [a Happy New Year, a safe journey, *etc.*]**. On the whole, the entries in the *Second Interim Report on Collocations* were drawn from existing English learning materials, word lists and dictionaries (most notably Saito[2], 1915) and re-classified by Palmer and his colleagues to show their basic phraseology and syntactic patterning.

In order to select collocations, Palmer took the "degree of intelligibility" (1931, p.18) as the main criterion. He interpreted collocation as applying to "all successions of words likely to be unintelligible or only partly intelligible to the foreign student even when he has a fair knowledge of each of the component words of the succession"

(Palmer, 1933a, p.1). Here, he distinguished between:

- "abnormal collocations" (i.e., the meaning of the collocation cannot be deduced from its parts);
- "semi-normal collocations" (i.e., collocations that can be immediately understood if the more particular senses of the component words are known);
- "normal collocations" (i.e., immediately comprehensible combinations).

The list of collocations included "all abnormal collocations of a given radius" and excluded "all normal collocations" (Palmer, 1933b, p.21) because the latter could not be limited in any way. Semi-normal collocations were included if they belonged to very high frequency radii, but normal collocations that could be easily substituted such as **want someone to do something** or **some time ago** were omitted.

Palmer's introductions to the two reports not only explain the careful decision-making in collocation selection, but also present wider views about why such an overarching lexical approach was necessary. Two points in particular stand out. The first concerns what we might now call lexical sets or clusters. The second addresses how lexical knowledge may underpin language proficiency. Both touch on associative links. Palmer argued that learners do not deal with isolated words when they learn vocabulary of a language. He believed that words are linked by 'thought associations' into constellations of words (Palmer, 1931, pp.32-33). These constellations include individual items that have both collocation and associative links (see Appendix 1.1 for an example constellation). Palmer's notion of thought associations led him to claim that language proficiency was "little other than proficiency in the extemporizing or the understanding of such normal comings-together of words" (Palmer, 1931, p.19)—a view which clearly foreshadows later ideas about formulaic language (e.g., Nattinger & DeCarrico, 1992; Wray, 2002) and probable connections between L2 collocation knowledge and fluency (e.g., Pawley & Syder, 1983; Lewis,

2000; Teubert, 2004).

Some of Palmer's ideas were directly derived from de Saussure and others. It appears that Palmer based his three-way classification of collocations on Bally's analysis of the French language. According to Haussman, Bally differentiated "les associations libres et occasionelles, les séries phraséologiques ou groupments usuels et les unités indissolubles" (Haussmann, 1979, p.189) in developing various collocation-based French learning materials for foreign students at Geneva University in 1909—some 20 years before Palmer's reports. Bally was also a student of de Saussure's who, with Sèchehaye, compiled the *Course in General Linguistics* (de Saussure, 1915) from de Saussure's lecture notes after the latter's death. As Palmer was in written correspondence with Sèchehaye about de Saussure's ideas (Hornby, 1946, p.8), we may assume that Palmer would have also been in contact with Bally or at least known of his work on phraseological groups. Equally, de Saussure's theory of language contained certain elements that Palmer appropriated. Part of de Saussure's view of the mechanism of language involves syntagmatic and associative relations. According to de Saussure, language use creates syntagmatic solidarities between two or more sequential units, whereas individual speakers create their own associations that are not fixed in order or number (de Saussure, 1911, pp.124-126). Yet, neither de Saussure nor Bally used the term collocation, so one of Palmer's fundamental contributions was to bring previous theory into a decidedly practical framework and to baptize it formally.

In sum, Palmer's work on collocation was pedagogically driven and exhaustively taxonomic. He used intelligibility as the principal measure by which to classify collocations into three major phraseological types. He also suggested potential connections between associative networks, collocations and fluency, but did not position these ideas within a wider theory of language use. Quite possibly, the lack of

generalizable theory in Palmer's pragmatic interpretation of collocation distanced his ideas from the rapidly growing field of linguistics.

However, his close associate, Hornby, continued the work begun on collocation in preparing dictionaries for English learners. Hornby later collaborated with Cowie, an important figure in contemporary English collocation studies who has tried to restore Palmer's legacy (Cowie, 1999) and has himself made wider connections between the phraseological and the typological interpretations of collocation (Cowie, 1998a; 1998b). We will therefore continue this discussion of the phraseological approach by considering particular insights from the lexicographic work of Hornby and Cowie.

1.2.2 Building on the Palmer legacy: Hornby

Hornby worked with Palmer in Japan and developed further a Saussurian view of associative links in discussions of how learners might best learn vocabulary (Hornby, 1946). He distinguished between formal word families (such as **happy, unhappy, happily, unhappily, happiness, unhappiness**) and associative word families, i.e., "a group of words that enables the learner to form associations, associations of the sort that help him to think in the new language" (Hornby, 1947b, p.93). Such associative word families were seen by Hornby as important for helping the learner to "form the associations he needs if he is to acquire his new vocabulary in the right way" (Hornby, 1947b, p.95). An example associative word family is shown in Appendix 1.2. What is of interest is that these associative groupings have more obvious collocation links than the word constellations that Palmer had earlier proposed.

Like Palmer, Hornby was concerned with presenting collocations "to afford the user all possible help" (Hornby, 1947a, p.89) in understanding and producing them successfully in English. This concern was addressed through the preparation of learner

dictionaries (Hornby, Gatenby & Wakefield, 1942; 1948) where the results of Palmer and Hornby's research into collocations were practically applied (Smith, 1998, p.285). The lexicographic work led to an innovative codification of how fixed different collocations are, as shown below in Table 1.1 In the table, ten collocations from the *Second Interim Report on Collocations* (Palmer, 1933b) are compared with their presentation in the *Idiomatic and Syntactic English Dictionary* [ISED] (Hornby, Gatenby & Wakefield, 1942). Nine of the 10 collocations appear as examples in the dictionary, and three are marked in bold as idiomatic sub-entry examples.

Table 1.1 **Differential codification of example verb + object noun collocations (based on Cowie, 1999, p.60)**

The Second Interim Report on Collocations	*The Idiomatic and Syntactic English Dictionary*
to beat time	***beat time***
to break the law	*break the law*
to catch (a) cold	*catch (a) cold*
to catch fire	***catch fire***
to clap one's hands	--
to drop a hint	*drop a hint*
to earn one's living	*earn one's living*
to extend an invitation	*extend an invitation*
to find fault	***find fault (with)***
to follow someone's example	*follow someone's example*

In addition to this original categorization of collocation restriction, the ISED attempted to indicate what other words could be typically combined with a particular headword. This was done by bracketing possible alternatives to indicate other possible collocations (Cowie, 1999, p.64):

gentle *...as gentle nature [heart, look, voice, call, touch]*
love *...to love comfort [golf, playing tennis, sea-bathing].*

One other innovative feature of the ISED was the coding of the syntactic patterns that particular verb constructions took (Béjoint, 1994, p.66). This meant that a great deal of early work was completed on what would be later covered by research into colligation (Firth, 1956) and the specific patternings of individual verbs (e.g., Hunston & Francis, 2000). All in all, Hornby's engagement with helping learners to recognize and produce collocations involved dealing with several key questions of phraseology. In the British lexicographic field, it was Cowie in particular who extended Hornby's phraseological interpretation of collocation.

1.2.3 Extending the phraseological interpretation: Cowie

Cowie worked under Hornby on the third edition of *The Oxford Advanced Learner's Dictionary* (OUP, 1974) and was co-editor of *The Oxford Dictionary of Current Idiomatic English* [ODCIE] (Cowie & Mackin, 1975). He later became editor-in-chief of the fourth edition of *The Oxford Advanced Learner's Dictionary* [OALD4] (OUP, 1989). Although Béjoint singles out the OALD4 for "the particular care taken in the indication of collocations" (Béjoint, 1994, p.67), it is in the extensive introduction to the ODCIE and two papers prior to the publication of the OALD4 that Cowie set up his collocation stall (Cowie, 1978, 1981).

Explicitly geared to "the practical needs of the learner" (Cowie & Mackin, 1975, pp.xiii-xv), the introduction to the ODCIE also explained at some length how collocations were organized (pp.lxii-lxvi). The dictionary was very much intended to bring together "associated words in one place" for headwords (p.xiv) so that the process of learning collocations could be accelerated and the learner might create their own collocations. Information about grammatical patterns and collocation ranges was also

provided in the way that the entries were organized. Semantic sub-classes in collocation ranges were, for example, made explicit by separating them with a semi-colon. In the following example (Cowie, 1978, p.136), **O** indicates the collocation functions as object, and **S** as subject:

> **hold up** ...delay; halt ...**O**: production, delivery; traffic, travellers.
> **shoot up** ...rise, increase, sharply...**S**: price, cost, rent; temperature, pressure; applications, attendance.

The treatment was more sensitive than the schema used by Hornby in the ISED. The other important point here is that the example collocations were intended to be representative and suggestive, so that the learner would be encouraged to interpret and "work directly with the lexical material, even though this may be limited" (Cowie, 1978, p.136). In particular, an explicit difference was made between open and restricted collocations, with restricted collocation sets marked off by a triangular warning sign: "The foreign learner would be wise to regard this sign as a warning to confine himself for some time to the choices indicated" (Cowie & Mackin, 1975, p.lxvi). This authoritative pedagogic guidance was very much in keeping with the down-to-earth, pragmatic approach of Palmer and Hornby—as though the lexicographer were more classroom teacher than theoretician.

However, the late 1970s and the 1980s saw the outbreak of the so-called "dictionary war" (Béjoint, 1994, pp.77-79) for market share between publishers of learner dictionaries. As if in response to these competitive and de-centering forces, Cowie moved towards greater theorization of collocation, particularly with regard to **collocation restriction**. The pragmatically simple definition of collocation as "the co-occurrence of two or more lexical items as realizations of structural elements within a given syntactic pattern" (Cowie, 1978, p.132) now gave way to a focus on "composite units" (Cowie, 1981, p.225) that could be differentiated according to whether they were

open, i.e., "maximally variable" (Cowie, 1981, p.226) or subject to "extreme restriction" (Cowie, 1981, p.226), and whether they were figurative or literal. A further concern was how to differentiate collocation from idiom—a question that both Palmer and Hornby had also addressed, but which was now being interpreted not just by Cowie, but by a growing number of other researchers also. Because the answers lay partly in a typological interpretation of collocation, we will consider them in greater detail in the next section. Before that, we will briefly summarize the phraseological interpretation of collocation.

1.2.4 Summing up the phraseological interpretation

It is difficult to estimate the true quality of Palmer, Hornby and Cowie's achievement. On the one hand, so much other work has been accomplished in the field of collocation since the late 1970s. On the other, an historical evaluation, long obscured, is nevertheless difficult to come to because the field is still fast expanding. However, it is clear that the approach of these three pioneers was painstakingly careful, not just in the dogged pursuit of a better understanding of collocation, but also in the persistent quest to help learners of English to improve their collocation ability—at first, in Japan, and then later, across the world, with the global growth of ELT. We have seen how all three scholars tended to downplay the abstract theorization of collocation; we have also noted how Cowie, more than Palmer and Hornby, began to place collocation within a broader theoretical framework from the early 1980s onwards. Indeed, Cowie made connections to work done in the Soviet Union (Cowie, 1981, p.225), and this was the start of the bridge from the phraseological to the typological interpretation of collocation—which we turn our attention to next.

1.3 The typological interpretation of collocation

We now move to the other side of the bridge between the phraseological approach and the combinatory view of collocation. The latter approach is characterized here as typological in order to emphasize its formal, de-contextualized classification of collocations. As mentioned earlier, this interpretation has its early roots in the phraseological work of Vinogradov and Amosova from the 1940s to the 1960s—"the Russian tradition" in Cowie's terms (Cowie, 1998a, p.4). The typological interpretation of collocation also includes Western discussions of restrictions on lexical combinability in the 1980s and 1990s (e.g., Aisenstadt 1981; Benson, 1989; Cowie & Howarth, 1996; 1998) and the semantic labelling of lexical collocations (Benson, Benson & Ilson, 1986a; 1986b) based on the work of Mel'čuk and others on Explanatory and Combinatory Dictionaries or ECDs (Mel'čuk, 1998). We will take these three areas as our main reference points in considering the typological interpretation of collocation.

1.3.1 The Russian tradition

There is a bewildering diversity of terminology in Russian work on collocation (Strässler, 1982, p.23; Cowie, 1998a, pp.4-8), so we will limit our attention here to Vinogradov's theory of phraseological fusions, unities and combinations (Vinogradov, 1947) and the subsequent development of this tripartite categorization by Asomova (Asomova, 1963). We will refer to English summaries of their work as the original studies were written in Russian.

Vinogradov's work is widely considered as the cornerstone of the Russian view of collocation. His analysis was based on the motivatedness of combinations such as V+N phrases (Strässler, 1982, p.24; Cowie, 1998a, p.5). A **phraseological fusion** is

classified as unmotivated in that there is no discernible link "between the meaning of the whole combination and those of its components" (Cowie, 1998b, p.214), so that the phrase is semantically opaque and structurally frozen. A fusion is, it might be argued, similar to conventional views of the semantic opacity of idioms; it also bears a resemblance to the category of **abnormal collocation** that Palmer proposed. The second category in Vinogradov's schema, **phraseological unity**, involves partial motivation whereby one part of the combination figuratively extends the meaning of the complete phrase. Thus, **hit the hay** (Strässler, 1982, p.24) may be considered a figurative extension of what was once a physical action in having a sleep. The third category is **phraseological combination**. This type is completely motivated and involves "one component used in its direct meaning, while the other is used figuratively: **meet the demand, meet the necessity, meet the requirements**" (Cowie, 1998b, p.215).

Finding Vinogradov's three-way categorisation unclear, Asomova focused on re-casting the classification so that she could establish a "theory of collocations which establish a fixed context" (Strässler, 1982, p.25). Vinogradov's first two categories of phraseological fusion and phraseological unity were re-grouped under the single category of **idiom**, and the third category of phraseological combination was sub-divided into **phraseme** and **phraseoloid**. A phraseme refers to a combination where the restriction of figurative meaning applies to a single fixed word in a particular combination. For example, in the collocation **grind one's teeth**, the fixed restriction applies to the verb **grind** (Cowie, 1998b, p.215; Strässler, 1982, p.25). Phraseoloids, in contrast, do not have such constancy of restricted meaning. The term is used for collocations where there may be several combinable elements, each of which determines a different (and therefore not constant) meaning of the other element, as **in**

pay one's respects/ a compliment/ court to someone (Cowie, 1998b, p.215). Both Asomova and Vinograd try to pin down the semantic restrictions for collocations in a systematic way. Other researchers in the West later address the same problem, but arrive at somewhat different solutions.

1.3.2 Further restrictions on lexical combinations

Restrictedness can be seen in terms of the commutability restrictions between the twin components of V+N collocations, or the nature of selection restrictions on the company that words together keep together (Carter, 1998, pp. 56-57). To advance our understanding of the typological interpretation of collocation, we will briefly review four further attempts to classify commutability restrictions in V+N collocations.

Aisenstadt theorizes that **restricted collocations** (Aisenstadt, 1979) are "different from free collocations by their usage-restricted commutability" (Aisenstadt, 1981, p.54), with one or both elements of a restricted collocation limited in its commutability. He notes that V + (art) + (A) + N forms one of the main patterns for restricted collocations in English and argued that such collocations fall into three main types:

(i) the elements of the collocation have a "narrow and specific meaning", as in **shrug one's shoulders**, where the verb cannot combine with other body parts and the noun cannot be used with many other verbs;
(ii)one of the elements has a secondary, abstract meaning from its core sense, e.g., **pay attention**, where the verb in its figurative sense can collocate with a limited number of nouns such as **respects**, **compliments** or **condolences**;
(iii)the verb has become grammaticalised and lost much of its meaning (more commonly known as delexicalisation), as in **have a fall**. (Aisenstadt, 1981, p.58)

On the other hand, Benson (Benson, 1989, pp.4-5) presents a three-way classification for V+N combinations in terms of:

(i) free combinations;

(ii)combinations where the verb can be combined with a "large, but semantically limited number of direct objects", such as **run an airline, run an association** and **run a building society**
(iii)verb + noun combinations where the verb collocates with a limited group of nouns in a technical sense, such as **debug a program**. (Benson, 1989, p.5)

Benson's classification highlights near-idiomatic restrictedness, as well as the figuratively non-core and technically specialized senses that verbs can take.

The third classificatory scheme is offered by Cowie and Howarth who suggest four possibilities for grouping collocations by focusing solely on commutability restrictions. These are: (i) invariable collocations, (ii) collocations with limited choice at one point, (iii) collocation with limited choice at two points, and (iv) overlapping collocations (Cowie & Howarth, 1996, p.83).

In a later study, Howarth builds on this four-way commutability division and combines semantic elements from the Aisenstadt and Benson paradigms (Howarth, 1998). Howarth's proposition is that V+N collocations can be classified into five levels, which can then be finessed by a three-way semantic division of the verbs into technical, delexical and figurative senses. The five-level commutability restrictions are shown in Table 1.2, whereas Table 1.3 further below brings those restrictions together into the full 5x3 categorization. (The examples have been taken directly from Howarth.)

Table 1.2 **Howarth's criteria for classifying V+N combinations**

Level	Criteria	Verb features	Noun features	Examples
1	Freedom of substitution of the N; some restriction on the choice of V	A small number of synonymous Vs	An open set of Ns	*adopt/accept/agree to a proposal/ suggestion/ recommendation/ convention/plan/etc*
2	Some substitution of both elements	A small number of synonymous Vs	A small range of Ns can be used with the V in the given sense	*introduce/table/ bring forward a bill/an amendment*
3	Some substitution of the V; complete restriction on the choice of N	A small number of synonymous Vs	No other N can be used with the V in the given sense	*pay/take heed*
4	Complete restriction on the choice of V; some substitution of the N	There are no synonymous Vs	A small range of Ns can be used with the V in the given sense	*give the appearance/ impression*
5	Complete restriction on the choice of both elements	There are no synonymous Vs	No other N can be used with the V in the given sense	*curry favour*

Table 1.3 **Howarth's semantic division of verb types**

Level	Figurative verb sense	Delexical verb sense	Technical verb sense
1	*assume importance* *require qualifications*	*get satisfaction* *give evidence to*	~
2	*assume a role* *follow a procedure*	*give emphasis to* *have a chance to*	*carry a motion* *consider a bill*
3	*bring up children* *reach a conclusion*	*have access to* *make an application*	*bring an action* *receive Royal Assent*
4	*pay attention* *put sth. to use*	*do one's best* *take precautions*	*obtain a warrant* *publish a bill*
5	~	*make an investment have a bearing on*	~

Howarth's sophisticated model presents a 15-cell linguistic classification, but it may not be immediately clear how we can systematically categorise other V+N collocations within such a complex model.

Indeed, part of the problem with all the categorizations that we have reviewed here is the implied "continuum of formulaicity" (Wray, 2002, p.211) where the distinctions between one category and the next are not easily demarcated. In general, the classificatory schemes show how V+N combinations are formally constrained in terms of usage rather than use. None of the models that we have reviewed here takes account of context as a determining factor in collocation restriction. This is not surprising given the complexity of language use, but it does highlight the problems of formalizing collocations outside of specific context of situation and language use—of rank-shifting collocations, so to speak, so that formal generalizations can be made.

1.3.3 The semantic labelling of lexical collocations and lexical functions

The problem of categorizing collocations has been also addressed by lexicographers working on both theoretical and practical collocation dictionaries. Despite some earlier work done in the 1970s on the collocation environments of grammatical items (Sinclair, Jones & Daley, 1970) and the use of the term lexical collocations in the title of a related paper (Jones & Sinclair, 1974), the distinction between **lexical and grammatical collocations** was first made by Benson, Benson and Ilson (1986a, 1986b). This polar contrast deserves a brief explanation before we look at how lexical collocations have also been classified by lexical function.

BBI define a grammatical collocation as "a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause" (Benson, Benson & Ilson, 1986b, p.ix) and a lexical collocation as consisting

of "nouns, adjectives, verbs, and adverbs" (Benson, Benson & Ilson, 1986b, p.xxiv). They give eight types of grammatical collocation such as noun + preposition, noun + *to* + infinitive, adjective + preposition, and verbs (which are sub-divided into 19 verb patterns). Seven types of lexical collocation are presented, such as V+N, Adj+N, and V+Adv. BBI also use semantic macro-labels for classifying particular V+N collocations. These labels are for **CA collocations** (denoting creation and/or activation) and **EN collocations** (denoting eradication and/or nullification). Example CA collocations denoting creation include **compose music** and **reach a verdict**, while those indicating activation consist of such collocations as **set an alarm** and **launch a missile**; instances of EN collocations include **reverse a decision**, **repeal a law** and **ease tension** (Benson, Benson & Ilson, 1986b, pp.xxiv-xxvi). They describe EN collocations as "arbitrary and non-predictable" for which non-native speakers "must have a guide" (Benson, Benson & Ilson, 1986a, p.258) and explain that they have cautiously drawn on the work of Mel'čuk (among others) with lexical functions in developing an Explanatory and Combinatory Dictionary for Russian.

A recent account of lexical function analysis in ECDs may illuminate this final aspect of the typological interpretation of collocation (Mel'čuk, 1998). ECDs are theoretically-oriented and not intended for general purpose use. They are meant to provide detailed semantic and syntactic information about set phrases (or **phrasemes**). Mel'čuk observes that collocations "constitute the absolute majority of phrasemes and represent the main challenge of any theory of phraseology" (Mel'čuk, 1998, p.31). Using Meaning Text Theory focused on the spoken production of text, Mel'čuk puts forward a set of conditions by which phrases can be sub-categorized, and, ultimately, collocations described in terms of lexical function. At a deep structure level, a lexical function **f** represents a general abstract semantic notion that "associates with a specific

lexical unit [=LU], L, which is the 'argument', or 'keyword', of f" (Mel'čuk, 1998, p.32) and can be lexically realized in different surface forms. In these surface forms, according to Mel'čuk, two-word collocations such as **crack a joke** and **launch an attack** have one freely chosen constituent lexeme that retains its literal sense (in these examples, **joke** and **attack**), whereas the other component is restrictedly constructed because it is contingent on the first element. In the first example, the freely chosen 'argument' **joke** motivates the restricted selection of **crack** (pp.30-31). This is basically a deep structure approach to a semantic analysis of the lexical realizations and restrictedness of 'surface' collocations (see Appendix 1.3 for a sample lexical entry, ECD-style).

1.3.4 Summing up the typological interpretation

Mel'čuk's fascinating work illustrates one extreme of the typological interpretation of collocation. It is a hugely ambitious undertaking in what it seeks to include and explain; yet, despite this vast scope, it highlights in simple ways how the use of collocations is pragmatically motivated. The other classifications that we have briefly reviewed in this section attempt to formalize semantic restrictions in different ways. However, the wider recognition of textual effects and constraints on collocation is weak. Textually-derived concepts such as collocation range and cluster are outside the remit of the typological interpretation, as Carter notes: "The examination of collocational ranges of items begins where semantic analysis of selection restrictions leaves off" (Carter, 1998, p.57). We move then to the textual interpretation of collocation and the theories of Firth, Halliday and Sinclair in order to understand collocation in context.

1.4 The textual interpretation of collocation

The work of three major linguists—Firth, Halliday and Sinclair—informs the textual interpretation of collocation. It is no exaggeration to claim that their work has led corpus linguistics and, in particular, corpus-based collocation analysis.

1.4.1 Firth and collocation as a level of language

Firth was concerned with theorizing how meaning was produced at "mutually congruent series of levels" within language (Firth, 1957a, p.176), i.e., context of situation, collocation, syntax, phonology and phonetics. Although each level of the system was interdependent with the others, Firth was careful to distinguish **colligation** (Firth, 1956, p.113) within the syntactic level from collocation. Firth used the term colligation for the structural relationships between the formal categories (or elements) of the words used in particular collocations (such as Adj + N). These formal structural relationships were "abstractions from utterances" (Firth, 1957b, p.164), whereas statements about collocations applied to "mutually expectant orders of words" (Firth, 1957a, p.181) in restricted languages. It was, in not so many words, a probabilistic view of lexis that Halliday and Sinclair would later flesh out in greater detail.

By restricted language, Firth referred to language used in particular situations with specialized vocabulary and grammar (Firth, 1956, pp.112-113). Collocation was understood as "the study of key-words, pivotal words, leading words...in the company they usually keep—that is to say, an element of their meaning is indicated when their habitual word accompaniments are shown" (Firth, 1956, pp.106-107). For example, part of the meaning of the word **Berlin** in German came from it being often collocated in the 1950s with **Insel** (island), **Weltstadt** (world city) and **Congressstadt** (congress city) (Firth, 1956, p.113). Elsewhere, Firth states: "One of the meanings of

night is its collocability with **dark**, and of **dark**, of course, collocation with **night**" (Firth, 1957c, p.194). This view of collocation lies behind the declaration that "You shall know a word by the company it keeps!" (Firth, 1957a, p.179) and is dismissive of an essentialist semantic view where words in themselves have intrinsic core meanings. Rather, according to Firth, collocation is a central dimension in understanding how meaning and functional value are created through use: "The distribution of the collocations in larger texts will probably provide a basis for functional values or meanings for words of all types" (Firth, 1952/53, p.23).

Firth distinguished collocation from context and citation. Context was a higher order level of meaning at which "the whole conceptual meaning is implied" (Firth, 1957a, p.180), whereas citation had a lexicographic function in specifying meaning "in shifted terms" (Firth, 1957a, p.180). Shifted refers here to de-contextualized and generalized meaning beyond a specific context of situation (Firth, 1952/1953, p.20). It was, according to Firth, for the lexicographer to choose particular definitions of meanings for words from observing their habitual collocations. Thus, part of the meaning of **cows** is indicated by observing collocations such as **They are milking the cows** and **Cows give milk**. Other words such as **tigresses** and **lionesses** do not have such collocations so their meaning can be separated at the collocation level from that of **cows** (Firth, 1957a, p.180). Despite the artificiality of the example, the principle of sense distinction by collocation pointed both to a fundamental characteristic of language use and a more far-reaching view of collocation than the de-contextualized uni-dimensional taxonomies that had concerned Palmer.

Overall, Firth's theorization of collocation is notable for pre-figuring how corpus linguistics would later analyze and capture the senses of particular lexical items through examining their collocation environments. His other major contribution to the

development of collocation theory rests in identifying collocation as a level of system within language use and proposing that lexicography should proceed by collocation analysis of texts. This work prepared the stage for Halliday and Sinclair to look more closely at lexical relations in text and to show more precisely how collocation can be analysed as part of the level of lexis in language.

1.4.2 Firth's theory applied: Halliday's interpretation of collocation

Halliday's work on collocation lies somewhere between the wide-ranging claims of Firth and Sinclair's meticulous, data-driven development of a lexical understanding of language. In *Lexis as a Linguistic Level* (1966), Halliday discusses how "lexicalness" may be appropriately described. He suggests that lexical relations should be considered separately from grammar, because the grammatical relations that lexical items enter into are not in themselves necessary for a lexical analysis. Rather, it is the lexical patterns in which lexical items co-occur that are important. Thus, **a strong argument, he argued strongly, the strength of his argument** and **his argument was strengthened** all exhibit the same lexical relationship independently of the differences in their formal realization (Halliday, 1966, p.151).

To put things more precisely, Halliday posits, alongside the paradigmatic category of **set**, the syntagmatic category of **collocation** for understanding lexis in language. The intersection of these two axes allows analysis of "a very simple set of relations into which enter a large number of items" (Halliday, 1966, p.153). Collocation restricts the co-occurrence of particular lexical items and may allow for prediction of items that co-occur "with a probability greater than chance" (Halliday, 1966, p.156). Further, if the lexical items in lexical sets are restricted to those that collocate with each other, then lexical sets can be delimited by a statistical procedure. It is, then, the co-occurrence of

lexical items that is the foundation for analyzing and interpreting the lexical level of language in specific ways. Halliday suggests that 20 million words of text would be necessary to decide "collocationally defined sets with citations to indicate the defining environments" of frequent lexical items in English (Halliday, 1966, p.160), but he does not pursue this argument for corpus-based analysis further.

In his later work (e.g., Halliday & Hasan, 1976; Halliday, 1985), Halliday narrows the focus to short texts of several sentences in order to analyse how both reiteration and collocation help create lexical cohesion and sustain text as discourse: "When we consider cohesion...we are investigating the linguistic means whereby a text is enabled to function as a single meaningful unit" (Halliday & Hasan, 1976, pp. 28-30). Lexical items which "are in some way associated with each other in the language" (Halliday & Hasan, 1976, p.285) have the potential to create lexical cohesion in text "if they occur in adjacent sentences" (p.286). Here, less frequent lexical items have a greater role in **collocational cohesion** (p.287) than highly frequent items. Although Halliday assigns a particular role for collocation cohesion in text, this is put forward as something of a catch-all category for everything not covered by reiteration. The collocation relations between related lexical items in text are not further differentiated. Halliday later claims that collocation rather than synonymy between related lexical items is more likely to create lexical cohesion in text (Halliday, 1985, p.313), but the claim is weakly developed.

Theoretically insightful in its early arguments, Halliday's work on collocation remains somewhat general in its later application. Following Firth, he claims collocation as a level of textual analysis. Yet, text takes on a markedly restricted meaning with the explicit focus on collocation links between neighbouring sentences. Moreover, the relatively few examples provide but passing glimpses of how collocation

permeates text. A more pervasive, data-driven interpretation is developed by Sinclair in his analysis of collocation in text.

1.4.3 Firth's theory applied: Sinclair's interpretation of collocation

An enduring concern in Sinclair's theory of collocation is how humans use language, keeping some parts fixed, others varied, yet, with a beguiling sense of agility, manage to maintain naturalness. If collocation does not necessarily always apply to adjacent lexical items, then how can the variable environments of lexical items across different texts be measured by a fixed criterion? Sinclair's early work was concerned with delimiting the basic features of a model so that such questions could be addressed. He proposed the **node** as the fixed variable (i.e., the lexical item whose collocation environments are under examination) and argued that the **span** of the node (the number of other lexical items co-occurring before and after the node) should be varied until it could be fixed "at the optimum value" (Sinclair, 1966, p.415). Such an analysis allowed a **cluster** to be identified (i.e., a set of typical collocates derived from all the collocates that co-occur with the node), where the **range** of a cluster would indicate the probability of items within a cluster inter-collocating with each other. For example, **vote** would be more likely to co-occur with more lexical items and thus have a greater range than **poll**. Drawing an important difference between **casual collocation** (i.e., chance) and **significant collocation** (i.e., typical and repeated), Sinclair proposed a basic formula for predicting the probability of two lexical items collocating with each other in a particular text. Finally, he also discussed why, in collocation analysis, it might be necessary to conflate word forms of the same lemma. Further work would show whether this held or not (Sinclair, 1966, pp.419-425). This was altogether a tentative, but far-reaching, theoretical agenda that much subsequent research would test

and expand.

Two projects in particular stand out as landmarks in this text-driven development of collocation theory. First, a co-authored report on collocation research conducted between 1967 and 1969 (Sinclair, Jones, & Daley, 1970) shows how the optimum span was empirically defined as +/- four words right and left of the node. The report presents different statistical analyses of the collocation environments of highly frequent 'grammatical words' (such as prepositions and articles) and the collocation behaviour of particular lexical items. Of specific interest is the development of collocation networks that show empirically tracked relationships between nodes and collocates (see Appendix 1.4 for an example network). Such a complex web of interrelating real-world associations offers an interesting contrast to the rather idealized networks presented by Palmer and Hornby (see Appendices 1.1 and 1.2). Of particular originality is the discussion of **lexicality** (Sinclair, Jones, & Daley, 1970, p.4) and its inverse relation to frequency. The more frequent a lexical item, the less lexical it will be, and vice versa. Highly frequent lexical items are "more grammatical/less lexical" and therefore bound to certain positions in collocation patterns, whereas less frequent items have greater positional freedom. The treatment of lexicality foreshadows later elaboration of **delexicalisation** (Sinclair, 1987a, p.323) and **collocation structure** (Sinclair, 1985, p.91).

The other ground-breaking project, the Collins Birmingham University International Language Database or COBUILD, involved the establishment of a 7.3 million word computerised corpus of English. This led to the production of the first learner dictionary based on automated, corpus-based concordance analysis of lexical items and their habitual collocation environments (HarperCollins, 1987). The COBUILD project yielded a wealth of insights into how common words in English

collocate with each other. First, it was found that the most frequent of words of English tend to be collocated in delexical senses rather than in a full lexical sense so that they "function as elements of structure" (Renouf, 1987, p.177). Second, the decision to include in the dictionary unaltered examples of real language use led to a growing awareness of how language users "use sentences to form longer structures, and so rely heavily on the language surrounding them for the full communication of meaning" (Fox, 1987, p.148). In other words, while the concordance analysis provided discrete evidence of collocations for distinguishing word senses, collocation itself was now becoming more clearly understood as a level of language use or "lexical realisation of the situational context" (Moon, 1987, p.92)—as Firth had originally claimed. For example, the differing textual collocates of **skate—ice**, **roller** and **winter** for sporting activity, and **fish**, **ray**, **shark**, and **water** for fish (Moon, 1987, pp.91-92)—reveal the distinct contextually bound meanings of the item. The third major insight was that the different senses of lexical items had such constrained typical phrasal patternings that few frequent words could be thought "to have a residue of patterning that can be used independently" (Sinclair, 1987c, p.158). This view of collocation started to dissolve the traditional division between grammar and lexis and would come to full fruition in later work on the phraseology of lexical grammar (e.g., Hunston & Francis, 2000). However, perhaps the most far-reaching insight from the COBUILD project was the overriding sense of "the replacement of words by phrases" (Sinclair, 1987c, p.150) in the real-world use of English. If the "normal use of English is to select more than one word at a time, and to blend such selections with each other" (Renouf & Sinclair, 1991, p.143), what exactly might this further entail for the textual interpretation of collocation?

In narrow terms, the repeated collocation of the everyday frequent words of

English leads towards their "progressive delexicalisation" as individual items (Sinclair, 1987a, p.321) and weakens the whole idea of individual words having independent core meanings. Here, an important difference is made between **downward collocation** and **upward collocation** (Sinclair, 1987a, pp.325-331). Downward collocation refers to the combination of a less frequent word as collocate for a more frequent word as node. An example is the collocation of **trace** with the node **back** in the sentence: **These could be *traced* back to the early sixties.** In contrast, upward collocation means the combination of a more frequent word as collocate for a less frequent word as node. A case in point involves the collocation of **to** with **back** in the same example sentence: **These could be traced back *to* the early sixties.** Downward collocation points to greater delexicalisation/grammaticalness, while upward collocation produces greater semantic value.

More broadly, Sinclair argued that two different principles of interpretation would be needed to explain adequately how multi-word units create meaning in text. The two principles are "the open choice principle" and "the idiom principle" (Sinclair, 1987a, pp. 318-319; 1991, pp. 109-121). The **open choice principle** refers to a traditional paradigmatic view of text construction that sees the process as a series of open-ended choices about what is to be slotted and filled in between completed units such as a word, phrase or clause. The **idiom principle**, on the other hand, refers to the syntagmatic axis and the contextually constrained set of "semi-preconstructed phrases that constitute single choices" (Sinclair, 1987a, p.320) that a speaker or writer can draw on to textualize meaning in ongoing text construction. It is the idiom principle, in Sinclair's view, that predominates and is "far more pervasive and elusive" (Sinclair, 1987a, p.321) than previously imagined.

To elucidate the idiom principle, Sinclair at first sees collocation and colligation as

having a singularly major role. An example may serve to illustrate this. The phrase **set eyes on** attracts into its environment "a pronoun subject, and either *never* or a temporal conjunction like *the moment*, *the first time*, and the word *had* as an auxiliary to *set*" (Sinclair, 1987a, p.321). The question of delexicalisation across text also remains a central concern in understanding the idiom principle. In adjacent collocations such as **scientific study, scientific analysis, general trend** and **general opinion** (Sinclair, 1994, p.24), the adjective tends to be delexicalised and carries no strong independent meaning of its own; rather it emphasizes part of the noun's meaning. This is all by way of showing that lexical structures larger than the unit of the word are needed for tracing the distribution of meaning across text, but the question remains as to how these units of meaning can be adequately observed and demonstrated.

The evidence for the idiom principle in action is produced by layering the textual analysis in terms of the **core, collocation, colligation, semantic preference** and **semantic prosody** (Sinclair, 2004b, p.141; 2004c, p.34). These five categories of co-selection are defined in Table 1.4 below. **Core** and **semantic prosody** are obligatory categories, whereas **collocation, colligation** and **semantic preference** are optional.

Table 1.4 A model for reconciling the syntagmatic and paradigmatic axes

Category	Definition
core	the evidence of the occurrence of the item as a whole (to 4 words left & right maximum)
collocation	the co-occurrence of lexical items
colligation	the co-occurrence of lexical items with grammatical choices
semantic preference	the co-occurrence of lexical items with semantic choices
semantic prosody	the attitudinal and pragmatic meaning of selected lexical items / the determiner of the meaning of the whole

Sinclair explains how these categories work in relation to the word **budge**. The analysis is based on 30 concordances with **budge** as the node (Sinclair, 2004b, pp.142-147). A summary is given in Table 1.5 below. In the table, bracketed numbers indicate the total number of occurrences of a particular feature of the data.

Table 1.5 **The syntagmatic-paradigmatic model applied to budge**

Category	Realization
core	• There are 30 concordances of ***budge*** for analysis.
collocation	• Left collocates include **refuse to** [9], **not** [8], **n't** [8], **double negative** [1], **determined not to** [1], **determined not to** [1], **has yet to** [1]. The other concordance focuses on a long and unpleasant period before the use of **budge**. • The immediate right collocates are preceded by particular punctuation marks in 15 cases—full stop [12], comma [2], dash [1]—indicating the use of **budge** as an intransitive verb. • Many right collocates are varied [10]. • There are several right collocate direct human objects [5], showing the use of **budge** as a transitive verb.
colligation	• These are mainly verbs (**refuse to / determine not to**) and modals of refusal or inability (**can't, won't, would not, could not**, and **didn't/did not**).
semantic preference	• Based on the collocation and colligation patterning, semantic preference is interpreted as expressing refusal in the intransitive verb patterning [25] and inability in the transitive verb patterning [5].
semantic prosody	• By looking at the wider co-text outside of the immediate span of the node, it becomes clear that the user wishes to express or report frustration (or a similar emotion) at the refusal or inability of some obstacle to move, despite pressure being applied [30].

The analysis shows the idiom principle mainly in action to the left of the node where the collocation and colligation patternings are more frequent. There are phrasal restrictions to the right in only five cases. Thus, the right-hand environment of **budge** shows a tendency towards the open-choice principle. Both principles can be thus observed in the analysis of this lexical item. The analysis of **budge** concentrates on the intersection of the paradigmatic and syntagmatic axes. It renews connection with a Firthian interpretation by approaching the data at different levels within that intersection. There is, as Sinclair acknowledges, enormous variety in the realization of the syntagmatic axis within a particular limited set of coordinated lexico-grammatical choices. Yet, consistent patterning in the production of meaning is nevertheless evident and open to analysis.

To conclude this presentation of Sinclair's 40-year engagement with a textual interpretation of collocation, the most striking aspect is the ongoing tentative testing of an evolving theory of collocation against the data of real language use, in all its messiness and apparent chaos. This allows for the development of more finely attuned analytical procedures for observing and explaining collocation and the natural texturing of meaning that is so effortlessly achieved by language users.

1.4.4 Summing up the textual interpretation

The textual interpretation of collocation is so much driven by corpus analysis that it is hard to imagine collocation as anything other than "a textual phenomenon" (Hoey, 1991, p.219). This phenomenon has been most widely pursued by the Birmingham University group, both past and present, which, brought together by Sinclair in the 1980s, has explored all manner of collocation issues through using corpus evidence. Other important work from 'The Birmingham School' includes Hoey (1991, 2001, 2005) for

developing theories of how collocation works across complete texts, and not just in adjacent sentences; McCarthy (1998) for investigating collocation in spoken discourse; and Moon (1998) for exploring the idiomatic angle of fixed expressions in English. As the above presentation has shown, the pioneers of this wide-reaching enterprise were Firth, Halliday and Sinclair. It is with good reason that Sinclair has been described as "one of the major figures in world linguistics" (Carter, 2004, p.1), for the textual interpretation of collocation has in many ways transformed how we understand not just collocation, but language use in general.

1.5 Drawing things together: Two *is* company and three *is* a crowd, but...

We set out in this chapter to gain a sense of three distinct interpretations of collocation so that we could achieve two goals. First, we wanted to raise issues about researching collocation; second, we aimed to position the following investigation of the acquisition of second language collocation knowledge against this broader background.

The chapter opened with a question "Two's company, three's a crowd?", which has, I hoped, served to underline certain enduring tensions in how collocation has been researched and interpreted. At a very general level, it is not altogether possible to keep three distinct schools of interpretation—the phraseological, the typological and the textual—so neatly compartmentalized. We saw how the phraseological interpretation started to seep into the typological, pushing the 'classical' lexicographic tradition of Palmer, Hornby and Cowie into a closer alignment with a de-contextualized semantic classification of collocation. Such concerns arose from trying to deal with collocation in isolation in ever more specific types of collocation dictionary. We noted too how the textual interpretation of collocation came to pay closer and closer attention to the creation of meaning through collocation in discourse—as if collocation itself cannot be

so easily pinned down into the specific two- and three-word lexical combinations that the other two interpretations have emphasized so much. There appears, then, to be a huge gap between different approaches to interpreting collocation.

Although this chapter may have been crowded, dissimilar interpretations are nevertheless good company: Meaning is contrastively created, and we can only know each interpretation well by noting overlaps and also considering different, even opposite, views of the same phenomenon. Against this complicated and notably unresolved set of background issues, we will look in the next chapter at how second language collocation knowledge has been researched and interpreted.

Notes

1. Palmer used the term **collocation** in a discussion of idioms in 1929: "Some say that an idiom is a collocation of words peculiar to any one language" (Palmer, 1929b, p.1). In the same article, Palmer wrote: "Others consider as idioms any groupings of words that cannot be understood by one who knows only the meaning of each of the component words" (Palmer, 1929b, p.1). This formulation is somewhat close to what Palmer would later use **collocation** for, but in 1929 he had not yet decided on the precise name. Rather, it seems that he was considering coining a new term with the root **~log**: "...so might lexicologists work out their terminology on derivatives of the word **log** instead of the almost meaningless term **word**" (Palmer, 1929a, p.2).

2. Palmer notes in a 1934 report for the Institute for Research in English Teaching (IRET): "In 1930 we produced our first tentative selection of 3000 words...On the same occasion we presented in mimeographed form a rough draft of a collection of collocations (culled for the most part from Saito's *Idiomological Dictionary*). This constituted our *First Interim Report on Collocations"* (Palmer, 1934, p.20).

Chapter 2
Literature Review of Second Language Collocation Research

2.0 Overview

Despite the wealth of L1 theoretical work and research into collocations that we noted in Chapter 1, the cupboard is somewhat bare when we look to L2 collocation research. Many accounts tend to be pedagogic, and the actual number of research studies of L2 collocation knowledge and use is limited. Fifteen such studies could be identified for the period 1989-2003. The limited body of previous research makes it difficult to identify clear progressions from one study to the next. The research questions tend to be specific to the particular study, and, without a common research agenda, there is a restricted sense of interconnection. The studies are mainly one-off, isolated single interventions. This lack of explicit continuity suggests that a chronological review would be both fragmented and disconnected. However, the previous work can be classified into four groups by the type of research approach taken, as shown in Table 2.1. As several studies use multiple instruments to measure various aspects of L2 collocation knowledge, the four groups in the table are organised according to the main research method.

Table 2.1 Classification of previous L2 collocation studies

Group & main focus	Author(s)
Group 1: Phrase- & sentence-level analysis	Biskup, 1992 Bahns & Eldaw, 1993 Farghal & Obiedat, 1995 Bonk, 2000
Group 2: Prompted essay with large-scale collocation inventory analysis	Zhang, 1993 Gitsaki, 1996
Group 3: Essay corpus with narrow analysis of specific collocation types	Dechert & Lennon, 1989 Chi, Wong & Wong 1994 Granger, 1998c Howarth, 1998 Nesselhauf, 2003
Group 4: Experimental measures of collocation knowledge	Read, 1993 Read, 1998 Schmitt, 1999 Mochizuki, 2002

The four-way categorisation forms the basis of the literature review in this chapter. Each group of studies is introduced with a brief explanation of the common methodological approach, and then each individual study in a group is summarised and evaluated in chronological order. A short commentary concludes the presentation of each set of studies. From this review, I then identify relevant issues for investigating L2 collocation knowledge further and conclude this chapter with a brief discussion of those issues.

2.1 Phrase- & sentence-level analysis

The first group of papers is comprised of four studies that investigate L2 collocation knowledge predominantly through phrase- and sentence-level analysis. Phrase- and sentence-level analysis here means a focus on (a) particular collocation types presented out of context, such as isolated V+N combinations, and (b) particular collocations presented in single sentences. Translation of collocations, sentence-level cloze, and matching target collocations to sentence-level paraphrases are the main instruments used.

2.1.1 Study 1: Biskup (1992)

Summary: The study aimed to compare two groups of learners with different L1s (Polish and German) in translating a set of collocations from their respective mother tongue into English, as well as investigate the causes of any interference in their L2 productions. The two groups consisted of 34 Polish English majors and 28 German English majors, with 10 years' experience of learning English, who are characterized as "very advanced" (p.88). Both groups were required to translate lexical collocations into English, with their responses later evaluated by three English NS on a 4-point scale from "unacceptable" to "full equivalent." Although Biskup does not explain how many lexical collocations were tested, it appears, from dividing the total number of responses presented in Table 8.2 (p.88), that there were 23 items in total (i.e., 782/34 and 644/28). We do not know for certain whether these 23 lexical collocations were presented at

the sentence level or as isolated phrases. It is also unclear what types of lexical collocations were used in the study, although the later discussion of the results seems to indicate that both **to** + V + N and Adj + N items were among those included.

The results showed that both groups produced the same mean ratio of correct responses (23%). The Polish group produced more restricted collocations than the German group (5.21% vs. 3.82%), and the German students tended to supply paraphrases more than the Polish (16.43% vs. 13.65%). On the qualitative front, two main points are noted. Firstly, the German group tended to produce synonymous free combinations if they did not know a particular restricted combination, whereas the Polish group tended to avoid supplying incorrect answers. Secondly, there was little overlap in the English translations of V + N collocations between the two groups.

Looking at the incorrect variants more closely, Biskup claims that the results for the Polish group indicated L1 semantic interference/transfer, whereas the German students produced erroneous forms based on "assumed formal similarity" (p.91). She attributes this difference to the closer linguistic proximity between German and English, and to the relative distance between English and Polish. Biskup briefly raises questions about L2 associative networks as being potentially useful for teaching L2 collocations, before concluding with comments about "semantic coreness" and "metaphor" as important parameters along which different language groups may organise sets of interrelated collocations.

Commentary: I will consider here the following two points: the claim that collocations are always transparent, and the contradictory ways in which results are displayed and conclusions drawn. First, in referring to an earlier pilot study (Gabryś-Biskup, 1990) into collocation perception and production, Biskup states that collocation recognition is unproblematic, whereas collocation production is not:

> As far as perception was concerned there was no visible difficulty for the learners since collocations are fully transparent, and comprehension tests produced 100 per cent correct answers. (p.86)

The generalization is somewhat sweeping, and some of the examples in this study tend to suggest otherwise. For example, the V + N collocation **run a bookshop** features the verb in its 5th most common sense (Cobuild, 1997) and its 15th combinatory pattern (Benson, Benson & Ilson, 1997, p.286), indicating that it can only be considered transparent if learners have developed their lexical knowledge beyond more frequent core senses towards abstract and metaphorical uses. This may be true of 'very advanced learners', but we cannot be sure what this descriptor means with regard to lexical knowledge. Biskup's study suggests that we need to be cautious in overestimating the ease with which different collocations can be recognised.

The second gap concerns the explicit presentation of data. This is weak. It is far from clear how many items were presented to the subjects, how these items were presented, and how they were divided between different types of lexical collocation. Although Biskup presents the seven types of lexical collocation listed in Benson (1985), we know neither the distribution of these types in the translation test nor how each type was realised. In effect, it is rather unclear what the learners were asked to translate, although we do get some indication later when Biskup discusses specific examples and their variants (pp.89-90). However, there is no specific account of which particular lexical collocations were more or less difficult for the two groups and individuals within each group. It is also unlikely that the differences between the groups were significant, which casts doubt on whether the later explanations of linguistic proximity can stand up to closer scrutiny. Biskup claims:

> In cases of verb + noun collocation, the translations of Polish and German learners seldom overlap, e.g. *to run a bookshop* rendered as *to drive a bookshop, to carry a bookshop, to introduce a bookshop* by Polish learners; *to manage a bookshop, to keep a bookshop* by German learners. (p.89),

but then presents contrasting V + N collocation translations from the two groups, where four of the student productions (**to make a record, to lead a bookshop, to measure s.o.'s pulse**

and **to feel s.o.'s pulse**) do coincide between the two groups. The effect of this contradictory presentation is to reduce the authority of the study.

Conclusion: Despite identifying V + N collocations as an important area for investigation, this paper falls short in several ways. Its most telling shortcomings are its generalised conceptual premise that collocation recognition is unproblematic and the incomplete and inconsistent way in which data are presented.

2.1.2 Study 2: Bahns & Eldaw (1993)

Summary: B&E investigated German English learners' productive knowledge of 15 V + N collocations through a translation task and a cloze task. The 58 students were 1st through 3rd year university students, with an advanced level of English and 7-9 years of prior English learning. The 15 collocations were taken from vocabulary learning materials and dictionaries, and the cloze test was piloted on two English NS. The English collocations were translated into German and put into sentences, and then the cloze task and translation task sentences were aligned. Thirty-four students completed the translation task, and 24 the cloze.

The student responses for the collocations were rated by three English NS as unacceptable or acceptable, after which B&E attempted to measure whether the students' collocation knowledge was equal to their general lexical knowledge. The total number of lexical words (83) in "the ideal translations of the German sentences" (p.105) was calculated by omitting prepositions, article and conjunctions. This yielded a total of 2822 lexical words (83 x 34), but was reduced to 2662 because of some blank responses. By looking at the differences between correct and incorrect responses for general lexical words and verb collocates, B&E observe that the subjects mistranslated 16.8% of the total lexical words, but nearly half of the inaccurately translated lexical words (48.2%) were verbal collocates. Of the verbal collocate total of 616, 35.1% were poorly translated. Moreover, the proportion of incorrect collocates did not significantly differ between the nine best and the nine worst translations (p.109, note 3). For the

cloze test, 48.1% of the responses had acceptable collocates, with 51.9% unacceptable. The mean of correct collocation responses for the cloze was 7.2, and 8.1 for the translation test, but no significant difference was found.

Through a more detailed examination of the collocation responses, B&E created two groups of phrases. The first group, scoring >85% for acceptable collocations in the translations, was comprised of **keep + diary, cancel + order, reject + proposal, do + damage, whip + cream**, and **achieve + perfection**. The second group, at <25% for acceptability for collocation translations, consisted of **serve + sentence, withdraw + money, refuse + admission, take + call** and **pay + compliments**. B&E comment:

> It would, of course, be very interesting and useful to know what differentiates one set from the other. The 15 collocations tested in our study are, however, too small a sample to venture any hypotheses on this point. (pp.107-8)

B&E were able to identify some V+N collocations that require explicit learning because they cannot be easily paraphrased. B&E conclude that L2 learners' collocation knowledge does not develop in parallel with general vocabulary knowledge. They comment that learners have better knowledge of general semantic than idiomatic restrictions and probably do not realise the difficulties of L2 collocation learning.

Commentary: Three points of specific interest arise: (a) differences in difficulty for particular lexical collocations, (b) estimating general vocabulary size, and (c) correlating vocabulary size and collocation knowledge.

Towards the end of their paper, B&E speculate as to the cause of difficulty in two sets of collocations. The first set tended to produce acceptable collocations in the learner responses:

keep + diary	**(Tagebuch führen)**
cancel + order	**(Bestellung rückgängig machen)**
reject + proposal	**(Vorschlag ablehnen)**
do + damage	**(Schaden anrichten)**
whip + cream	**(Sahne schlagen)**
achieve + perfection	**(Perfektion erreichen),**

but the second set led to relatively few acceptable collocations in translation:

serve + sentence	**(Haftstrafe abbüßen)**
withdraw + money	**(Geld abheben)**
refuse + admission	**(Zutritt verwehren)**
take + call	**(Anruf entgegennehmen)**
pay + compliments	**(Komplimente machen).**

In a solo paper from the same year (Bahns, 1993), Bahns suggests that the source of difficulty is governed by whether a particular combination can be directly translated into a learner's L1. This follows the argument that lexical non-congruence is a critical factor in collocation difficulty (Marton, 1977). However, it is not clear whether lexical congruence can cleanly separate the two sets of collocations above. **Tagebuch führen** and **keep a diary** do not map one-to-one onto each other, nor does **Schaden anrichten** with its full lexical verb correspond to the de-lexicalised verb **do** in **do damage**. Although the question of collocation difficulty is raised in this study, the answer seems to involve more than just one factor.

The second point of interest is B&E's estimation of general vocabulary size by counting the number of lexical words produced in total by the population. This result was generalized across all subjects, but individual distributions were not examined. There are other problems here. First, B&E used the same data source for estimating collocation knowledge and general lexical knowledge. The measures were not independent of each other, with different lexical items counted as both "lexical words" and "verbal collocates," so the internal validity is weak. Secondly, it is doubtful whether simply measuring the total number of tokens produced is adequate. It might be useful to distinguish type and token on a separate elicitation task, but an independent measure of vocabulary size needs to be considered. An interesting question is here whether ability in translation of lexical items should be taken as a reliable indicator of general lexical knowledge or of a particular type of lexical knowledge, namely facility in L1-L2 translation. Although B&E tried to correlate collocation knowledge to general vocabulary size, both types of knowledge needs to be more specifically operationalised and tested.

Conclusion: B&E use cloze and translation as indicators of productive collocation knowledge.

Such measures involve discrete items and push the emphasis towards accuracy as the standard of measurement, away from allowing for partial collocation knowledge. As a single intervention, any sense of collocation development is lost: the learner's knowledge is either total or not. Despite a number of weaknesses noted with the design, an important feature of B&E's study is the attempt to quantify both lexical and collocation knowledge.

2.1.3 Study 3: Farghal & Obiedat (1995)

Summary: F&O's study consists of two parts, an English language questionnaire (administered to 34 Jordanian university English majors = Group A) and an Arabic questionnaire (given to a separate group of 23 Jordanian English teachers with between 5 and 10 years' teaching experience = Group B). In the English part, the subjects were asked to supply the missing part of a collocation-pair in a gapped sentence already including one half of the pair (11 target collocations). An example sentence is: **To many people, cold food is better than _____ food**. The English version of the questionnaire was piloted on two English NS for a validity check. The Arabic part required subjects to translate the 11 collocations from Arabic into English. The 11 items consisted of one Noun + Noun collocation-pair (**depth of winter: height of summer**) and 10 Adjectival + Noun collocation-pairs (**calm sea : rough/stormy sea, hot food : cold food, light food : rich/fatty food, strong tea : weak tea, moderate drinker : heavy drinker, running color : fast color, salty soup : bland soup, strict rules : lenient rules, plain shirt : striped shirt**). Four of these 22 collocations are classified as "predictable collocations" (**calm sea, hot food, cold food**, and **salty soup**), i.e., directly translatable L1-L2 equivalences, with the other 18 termed "unpredictable collocations" i.e., L1-L2 divergent collocations (p.325; p.328, Appendix 1).

Data analysis revealed that the two groups used four lexical simplification strategies, namely, synonymy, avoidance, transfer, and paraphrasing, when they could not supply the correct collocation. The distribution of strategies was broadly similar for (i) avoidance (Group

A: 27%, Group B: 21.3%), (ii) transfer (Group A: 9.9%, Group B: 21.6%), and (iii) synonymy (Group A: 41%, Group B: 35.4%). Differences emerged for the use of paraphrasing (Group A: 3.8%, Group B: 25.1%) and correct collocations (Group A: 18.3%, Group B: 5.5%).

F&O explain the high frequency of synonymy as determined by lack of learner awareness of collocation restrictions for known everyday lexical items. Examples such as **oily food** and **greasy food** are given for **rich food**. The use of avoidance strategies resulted from learners choosing a "related natural collocation" (p.322) instead of retrieving the elicited collocation that they had previously met, e.g., **extravagant drinker** and **great drinker** for **heavy drinker**. Positive L1-L2 transfer occurred in cases of direct correspondence between Arabic and English (e.g., **striped shirt**), whereas the opposite held true for collocations without direct equivalence in the L2 (e.g., **heavy tea : strong tea, firm color : fast color**), suggesting that the subjects tended to over-generalise L1-L2 collocation equivalences in the latter cases.

The use of paraphrasing showed marked differences in the results between the two groups. F&O relate this difference to the constraining effect of the fill-in-the-gap English questionnaire used with Group A. Group B had more licence to paraphrase because they had to translate L1 collocations into English as best they could. Examples of paraphrase include **drinks too much** for **heavy drinker** and **does not change** for **fast color**. No explanation is offered for the differing results between the two groups for the use of correct collocations.

Commentary: I will limit my critical comments to inconsistencies in the selection of the collocation-pair, the assumed binary opposition between collocates, and the classification of predictable and unpredictable collocations.

It is not clear why F&O selected one N + N collocation-pair and 10 Adj + N collocation-pairs. It seems an oddly imbalanced division. F&O indicate that these are collocations that the subjects could be expected to have met in their previous English studies. However, we do not know whether these collocation-pairs are drawn from an officially prescribed vocabulary list or

set of textbooks used at university or secondary school, or if they were simply invented. Although this might not be such an issue if the items were commonly used collocations, the collocation-pairs contain several oddities: **light food (?): rich/fatty food, strong tea: weak tea, running color (?): fast color(?), salty soup: bland soup(?), strict rules: lenient rules, plain shirt: striped shirt.** Some of the gapped sentences also seem strange. For example, Item 5 for Group A is **To avoid gaining weight, most women eat light rather than _____ food**, and Item 8 is **This is a running color, but that is a _____ color**. We may have **a light meal**, but do we eat **light food**? Perhaps, we prefer **healthy food, low-calorie food**, or **simple food**. Similarly, **Some clothes are colourfast** and **The colour ran in the wash** both seem acceptable collocation predicative uses, but the attributive use of **running** and **fast** with **color** seems less appropriate.

Part of the problem is that the test items seem to be looking towards explicit binary opposition between collocates. In fact, several items may often form a collocation opposition within an existential paradigm (Brazil, 1997). Thus, for Item 3 (**To many people, cold food is better than _____ food**), possible real-world completions also include **no**, **warm**, **warmed up**, and **burnt** or even **badly cooked**. In the case of Item 6 (**I prefer _____ tea to strong tea**), **sweet** and **lemon** also seem to be possible responses. So, a number of weaknesses can be noted in the test instrument.

The final issue concerns the distinction made between "predictable collocations" and "unpredictable collocations," i.e., L1 and L2 collocations that do or do not directly converge. F&O note that **striped shirt** shows a "one-to-one correspondence between L1 and L2" (pp. 323-4), but this contradicts Appendix 1, which lists just 4 predictable collocations (**calm sea, hot food, cold food**, and **salty soup**).

Conclusion: This study throws light on the difficulty of testing for L2 collocation knowledge with a small number of items and a small population. The paper shows that selecting items

appropriately for a collocation test requires careful consideration, and F&O identify some basic strategies that learners may use in coping with collocation production.

2.1.4 Study 4: Bonk (2000)

Summary: Bonk targeted three types of collocation (verb-object collocations, verb-preposition combinations, and figurative-use-of-verb phrases) in a 50-item main test, with a shortened TOEFL test of 49 items used to estimate general English proficiency (N=98). Seventeen items in the collocation test involved supplying the missing verb in a V + N combination presented in sentence form (e.g., **I'm sorry, but I can't give it to you at that price. I'm trying to _____ a business here, not a charity organisation**). Another 17 items focussed on missing particles that had to be supplied after a verb in a sentence. Here, a prompt was given each time in the form of **to + verb**, which served as a near synonymous prompt for the partially deleted verb. For example, with **Many of the birds in this area were killed _____ by local hunters. (to exterminate)**, learners were expected to supply **killed off**. The final part of the test consisted of 16 items, each made up of four sentences using a particular verb in a different way, where the subjects had to choose the least acceptable sentence, for example:

a. Let's drop the subject. I don't want to talk about it any more.
b. He kept dropping hints that he wanted a job, too.
c. Bob said that he would bring the computer over to my house, but I dropped his offer.
d. She didn't like the professor, so she decided to drop the class.

Results showed that the V + Prep section had low reliability (K-R 20 = .47), with the V + N combinations scoring .69 and the figurative verbs .61. Overall, the test had a reliability co-efficient of .83. The shortened TOEFL test proved to be an adequate measure of general English ability (K-R 20 = .85). Ten out of 17 items in the V + Prep combinations had item-total correlations <.29, of which 8 were < .19. The Fig V section produced 6 items with item-total correlations < .29, of which 2 were < .19, whereas the V + N part yielded 5 weak items at < .29,

of which 2 scored <.19. The correlation between the collocation tests and the proficiency measure was .61, which, corrected for attenuation, decreased to $r^2 = .53$. Rasch analysis showed the V + Prep section to be weak, but overall "the test of collocations seems to have performed reasonably well" (p.20). Finally, factor analysis produced two factors, one related to knowledge of lexical relations, and the other to general language proficiency.

Bonk concludes that the test could be improved by replacing poorly discriminating items. The V + N test performed reliably, and qualitative analysis of 25% of the answer sheets showed that for the most part the subjects understood the prompt and the general schema targeted (e.g., ***break one's appetite** and ***destroy one's appetite** for **spoil one's appetite**), even if they could not supply the right collocation. Bonk notes a similar effect with the V + Prep section, but concludes that "the prepositions do not seem to be an extremely useful type of item to include in a test of collocations" (p.27). The V + Prep subtest was the least difficult, and the Fig V subtest the hardest, with the V + N subtest between the two. Bonk is cautious here because of the different formats for each of the three subtests and warns against seeing proficiency itself as an "extremely effective predictor of collocational proficiency" (p.29); rather, he argues for individual variation in collocation knowledge. He also hypothesizes that "well-developed collocations knowledge may be one of the last stages of second language acquisition" (p.30), but observes that low-level learners have some limited collocation knowledge.

Commentary: I would like to look critically at the design of the test and the question of collocation relationships. The instrument aims to test a large number of items, but the division into three sub-tests weakens this wider focus. It is also not clear exactly how the V + N, V + Prep and Fig V items have been selected. This lack of clarity can be highlighted by isolating the full set of items, as shown in Table 2.2 below. In Table 2.2, two contrasting examples involving **have** and **take** point to the first weakness in item selection—the uneven distribution of verb forms for the same lemma. **Have** comes up only twice in the V + N combinations (**have a**

dream, have the feeling), but **take** is included in all three types (**take pictures, take a nap, take a decision; take after; take off, take place, take turns, *take heart attack**).

Table 2.2 Target items (Bonk, 2000)

V + N	run a business, avoid the subject, take pictures, ruin your appetite, run the risk, have a dream, pay attention, make a difference, hold a meeting, keep a diary, take / have a nap, take / make a decision, have the feeling, do a favour, make a spectacle, wipe your feet, go out of business
V + Prep	come to, depend on, get over, set off, hold up, come out, give up, break off, kill off, look over, move on, pick up, pick on, take after, talk out, cheer up
Fig V	cover: with a gun / insurance / distance / *diagnosis take: ~ off / plane, ~ place / robbery, ~ turns / children, *heart attack put: blame / signature, put to death / murderers / *risk see: ~ off / person travelling, ~ to sth. / double / *~ up / plans leave: apartment / a possession to somebody / a spouse / *mind throw: party / game / into confusion / *concept drop: subject / hints / class / *offer stand: sth. negative / chance/ on an issue / *flowers pick: fights / pockets / berries / *conclusion play: radio / part / movie / *impact get: results / cancer / to a place / *success make: a good present / table / a total / make it (complete an action) do: repairs / best / well / *profession run: test / story / ~ over (in a car) / *vitamins move: car / be moved (by a film) / change homes / *situation pull: gun / weight / something (act in a surprising way) / *trick

A second weakness is the overlap between the categories of V + Prep and Fig V. One example is **take off, see off, and run over**, which are categorised as figurative but could have been also classified as V + Prep. In fact, many, even all, of the V + Prep items could be re-classified as a figurative use of the base verb. A third area of confusion is whether some of the examples are particle or preposition-based. Finally, we may note that the four items for **make** as a Fig V do not seem to include an incorrect 'figurative' usage, and that **~ a decision** in the V + N type suggests two possible choices, i.e., **make** and **take**. So, there are some weaknesses in item selection and classification.

My second concern is whether the test instrument measures collocation relationships. The effect of the cloze test is to limit the choice of relationship to a single possible link between the noun and deleted verb in the V + N sub-test, and the verb and deleted particle in the V + Prep

sub-test. In the latter case, the use of a semantically equivalent verb as prompt, e.g., **It's taken me more than a month to get _____ this cold. (to recover)**, makes this into a prompted recall test of synonymy, where the learner can either produce or not the required collocation. This is a limited measure of collocation production, and it is open to doubt what the results actually show.

Conclusion: The test does not shed light on how learners organise collocation relationships. It also suffers from confusion between the three organising types and from inconsistencies in item choice. The study does however question the widely accepted position that L2 collocation knowledge is an ability that advanced learners only develop and, by doing so, encourages us to question how less proficient learners might begin to develop their L2 collocation knowledge.

2.1.5 Concluding comments on the phrase- & sentence-level analysis studies

The four studies bear several similarities in the measurement of L2 collocation knowledge. The first common point is the emphasis on formal accuracy in translating particular items or supplying the missing part of a particular collocation. This leads to a concern with L1-L2 lexical congruence/equivalence and to the assumption that collocation difficulty partly rests in predictability. A second similarity is the small number of items tested; one effect is that only limited insights into L2 collocation knowledge are gained, although some quite large claims are made on the basis of such limited evidence. A third similarity is how the test instruments influence the conceptualisation of L2 collocation knowledge, which is, for the most part, seen as an advanced level ability to produce appropriate collocates. Collocation recognition, on the other hand, is claimed to be unproblematic. A further shared characteristic is the mixed selection of different types of collocation. Only Bahns and Eldaw's study limits its focus to one particular form, V + N collocations; the other three studies try to measure knowledge of different types of collocation, but for no consistent theoretical reason. Finally, two of the studies try to correlate collocation knowledge to other linguistic indicators, i.e., general lexical knowledge

(Bahns & Eldaw) and overall language proficiency (Bonk), but arrive at contradictory findings as to whether L2 collocation knowledge develops in parallel to overall (lexical) proficiency or is delayed until an advanced ability has already been established. In sum, these studies can be described as formally focused, small-scale single interventions which, despite some research design weaknesses, raise several interesting questions.

2.2 Prompted essay with large-scale collocation inventory analysis

The second group of studies consists of two large investigations of L2 collocation knowledge in which a corpus of essays is analysed for a large number of different types of grammatical and lexical collocations.

2.2.1 Study 5: Zhang (1993)

Summary: Zhang used three instruments in his study: an essay corpus written by NS and NNS first-year students, a collocation cloze test of 50 sentence-length items, and a context of use judgment test of the same completed 50 sentences as to whether they belonged to writing, daily speech, or both. The essay task asked the subjects to write about the benefits of certain types of recreation. After holistic rating into good, intermediate, and poor, 30 NS and 30 NNS essays were chosen for analysis (from a total of 83 essays); 15 'good' writers and 15 'poor' writers were chosen from either group, so that within-group and between-group comparisons could be made for grammatical and lexical collocations. The 50 sentences used in the two collocation tests were also piloted and revised before the main test administration. The collocation cloze items were analysed and scored according to a 4-point scale (p.80):

3 points	Formulaic, semantic, and grammatical
2 points	Not formulaic, but semantic and grammatical
	Formulaic, grammatical, but not semantic
	Formulaic, semantic, but not grammatical
1 point	Not formulaic, marginally semantic, but grammatical
	Not formulaic, not grammatical, but semantic
0 point	Not formulaic, not semantic, but grammatical
	Not formulaic, not semantic, not grammatical
	Missing responses.

Responses in the second collocation test on contextual appropriacy were matched against those of 136 NS respondents and scored 1 or 0.

The two sentence-based collocation tests showed significant differences in collocation knowledge between good and poor writers in each group, as well as between the NS and the NNS writers. The differences were greater for collocation knowledge in terms of form and meaning than for judgment of appropriate context of use. One other important finding was that good NNS writers retrieved collocations significantly more quickly than poor NNS writers in the collocation knowledge test. With the essay data, no significant difference was found between the NS and NNS writers in quantity of lexical and grammatical collocations produced (p.120). Zhang analysed the grammatical collocations in 8 categories comprising a total of 66 types. NS writers produced 57 of the 66 types, while NNS writers produced 47 types and did not use 18 of the types produced by NS writers (p.126). Zhang characterizes this as a difference in ability to use "the more structurally demanding" collocations (p.126). However, the most important significant difference was that NNS writers made more errors than NS writers. As for lexical collocation production, overall the NNS group produced slightly more than the NS group, with two types (V + Prep and V + {Prep} + {Det} + N) higher for the NNS. Closer examination of the data suggested that this resulted from greater repetition of lexical collocations by the NNS. The NS generally displayed greater variety and showed greater use of 2-part and 3-part verbs. Variety and accuracy were thus the more consistent indicators of NS-NNS difference.

Within-group differences with the essay data showed that the NS poor writers tended to use lexical collocations more than grammatical collocations, which suggested a lack of structural complexity. Moreover, the AUX category (tense, voice, mood) was weaker with the NS poor writers (pp.139-145). The NNS good writers differed from the NNS poor writers in quantity, variety, and accuracy with both lexical and grammatical collocations, but the difference between the two NNS groups was only significant for collocation errors for both categories of

collocation.

Commentary: Zhang's investigation is the first large-scale investigation of collocations in the studies under review here. Though it is on a different level from the four previous studies in the amount of data collected and analysed, it shares certain similarities and raises some important methodological issues. Here I will consider (a) the research instruments used, (b) the analysis of the essays for collocations, and (c) the NS / NNS comparison.

Zhang uses a simple controlled essay prompt to collect data from both the NS and the NNS groups. He is also careful to pilot and adjust the essay question and to administer the writing task under timed conditions. These are all improvements in such data collection. Less clear, however, is the 4-point scale for evaluating the cloze test responses, where the three components of formulaic, semantic, and grammatical are re-combined in different ways for scoring 2, 1, and O points. The overlap between each part of the scale is striking, so we cannot be sure whether it achieves reliable differentiation or not. It is doubtful whether such a complicated scale is warranted for judging one-word responses from the students. An example item, **The topic is familiar to me**, underlines the weakness of the scale: Would **The topic is familiar *for* me** score 1, 2 or 3 points?

The essay analysis involved a quantification of 66 types of collocation. For sure, this kind of approach aims to be comprehensive in its assessment of collocation use, but the net is cast so wide that it is unclear what the results show. Consider, for example, that an essay by student A includes 10 different collocation types, but they occur only once in every case. Another student, B, writes an essay that contains five different collocation types with a frequency of two, each one realised in a different way. We have the same quantity of collocations, and it would seem that we also have a similar degree of variety, but how are we to characterize the difference?

The third point of interest is the NS/NNS comparison. Zhang's results suggest that there are differences between poor and good writers in both groups. While I found this to be very

intriguing, I wanted to see specific examples of good and poor writing from either group.

Conclusion: The study by Zhang seeks to analyse comprehensively a large number of collocation types and to differentiate subjects by frequency of use of those 66 types. This pushes the focus towards a global assessment of written linguistic proficiency, but provides limited insight into collocation knowledge specifics, other than marking lexical error as a significant distinguishing factor.

2.2.2 Study 6: Gitsaki (1996)

Summary: Using 33 collocation types listed in Benson, Benson and Ilson (1986b), 2 adapted BBI types, and 2 taken from Zhang (1993), Gitsaki investigates whether stable patterns of collocation development can be identified within the same proficiency level and across different proficiency levels. Of the 37 types, 11 were lexical and 26 grammatical (see Appendix 2.1 for examples of each type). Three almost equal groups (N=275) of Greek learners of English (post-beginner, intermediate, and post-intermediate) were asked to complete a guided writing task, a translation exercise of 10 sentences into English (targeting 6 collocation types), and a sentence-level cloze test in which one part of a collocation was deleted (covering 12 collocation types). Gitsaki used holistic rating, lexical density, words per T-unit, S-nodes per T-unit, target-like use of articles, and error-free T-units to estimate English proficiency, all of which worked except for holistic rating. The groups were judged to belong to three distinct levels of proficiency.

The essays were quantitatively analysed for accurate use of the different types of collocation, with incorrect usages given 0 and the sum of correct usages per type by individual calculated. Results allowed the three proficiency levels to be characterized by the accurate use of the following types:

1. post-beginner group: SV c, Adjective Noun, and SV(O) Prep O;
2. intermediate group: Prep Noun, SV to Inf, Prep Det Noun, Phrasal Verb, Adjective Prep;
3. post-intermediate group: Noun Prep, SV Inf, SV (O) that, Adjective Prep.

Data from the translation and cloze tests were analysed by calculating the mean accurate response for each of the types targeted. Significant differences were found between groups. The post-intermediate group performed better for Noun preposition, SVV-ing, and Verb Noun (creation), but only outperformed the post-beginner group for the Adjective Prep type. The intermediate group outperformed both of the other groups for SV(O) Prep O. The Verb Noun (creation) collocation type turned out to be the most difficult to translate. The cloze test results indicated that the post-intermediate group was superior to the other two groups for all types, except where the post-beginner showed greater accuracy for Noun Noun and Adjective Prep. The post-intermediate group performed significantly better than the other two groups for eight of the 12 collocation types targeted.

Gitsaki concludes that it is possible to claim parallel development of collocation knowledge and language proficiency, with post-beginner learners first producing simple grammatical collocations and two-word lexical collocations unanalytically, albeit inaccurately. At the intermediate level, learners achieve greater accuracy with complex grammatical collocations, but still use lexical collocations inaccurately. The argument is that lexical collocations are first accurately acquired at the post-intermediate level, because they can be analysed by learners in both lexical and syntactic terms: the post-intermediate learner's ability to produce more accurate lexical collocations is thus dependent on a larger vocabulary size and greater colligation ability than at the intermediate level.

Commentary: One way of looking at Gitsaki's study is to see it as a modified research procedure from Zhang and to ask whether the problems noted in Zhang's study have been resolved. Simplicity is not a virtue in Gitsaki's study. It involves 3 independent variables, sorted against 37 dependent variables in the essay writing, 12 in the cloze test, and 6 in the translation test. Although the 37 collocation types in the essay task are filtered down to 14 distinct types across the three proficiency levels, one weakness is that the reliability checks of

translation and cloze coincide in only 6 of the total 37 collocation types. Very limited concurrent validity can thus be established between the different test instruments.

Because significant results occur across groups rather than individuals, things are further muddied as to whether a developmental sequence can be found to hold in particular productions or not. In the essay data analysis, text lengths of individual subjects are not standardized, so it is uncertain what the occurrence of a particular collocation type in a particular written production means. With unstandardized text lengths, it is also not clear what the frequency of occurrence of a particular collocation type points to. We might imagine that one essay is 100 words long, and another 200, and our analysis shows that they each have 10 different collocation types, but only 5 of the types are the same between the two essays. Should we take just those 5 same types for further analysis or should we take the 5 different types, or take all 10? Or should we reduce the 200-word essay to its first 100 words so that we can look at frequency and type with samples of the same length? Such variation will be a recurrent feature of the data, but it is difficult to account for. Finally, Gitsaki used holistic rating, lexical density, words per T-unit, S-nodes per T-unit, target-like use of articles, and error-free T-units to estimate the proficiency levels of the three groups. T-units work from clauses that can stand alone in a grammatically correct fashion, but it is by no means certain that formal accuracy is a robust indicator of proficiency or not. Perhaps more importantly, as we noted with Bahns and Eldaw, Gitsaki does not use an independent measure to establish proficiency.

Conclusion: The detailed scope of Gitsaki's analysis is impressive, but is somewhat unconvincing. A basic question is whether beginners produce lexical collocations more accurately than grammatical collocations, because they have memorized such 2-word collocations "as unanalysed blocks of language" (p.226). A counterview is possible: learners may combine individual words, from which some items become more strongly associated with each other and then later conventionally and accurately bonded as appropriate collocations.

2.2.3 Concluding comments on the large-scale collocation inventory analysis studies

Despite the increased data collection and the use of multiple instruments to analyse the data, these two studies remain single interventions aiming to produce comprehensive accounts of collocation production. This means that the broad claims in both studies are thinly spread across two rather exhaustive inventories. We noted that such analysis tends to obscure individual variation and development. As with the Bahns and Eldaw and Bonk studies, these two large-scale experiments also try to establish descriptions of L2 collocation knowledge in relation to different stages of overall linguistic proficiency; furthermore, Gitsaki proposes a broad model of the L2 lexicon in development. While these are obviously legitimate concerns for trying to understand L2 collocation development, the specifics are somewhat lacking. It remains, in other words, not clear how L2 combinations are initially organised in the L2 mental lexicon, how such organisation develops, and what influences L2 collocation production at different stages of development.

2.3 Essay corpus with narrow analysis of specific collocation types

The third group of studies uses essays as the main data source and focuses narrowly on specific collocation types. The types vary from study to study. Dechert and Lennon (1989) examine collocation blends, and Chi, Wong and Wong (1994) look at a small set of delexical verbs. Granger (1998c) analyzes the use of **–ly** amplifier collocates with a group of adjectivals and also considers how formulaic sentence builders are used. Howarth (1998) and Nesselhauf (2003), on the other hand, focus on the use of V + N collocations.

2.3.1 Study 7: Dechert & Lennon (1989)

Summary: D&L open their paper with this example of a collocation blend from a student essay: **Today, motoring offences are punished along similar laws in most European countries**. By collocation blend, they refer to the 'infelicitous' lexico-grammatical

combination of two habitual collocation syntagmas—in this case, **according to similar laws** and **along similar lines**. The research is based on a very small corpus of writing by German undergraduates who had studied English for 12 years and had advanced English proficiency. The corpus consisted of nine essays (p.166, note 1), and the study focuses on a close analysis of one essay written in response to a complex 4-sentence prompt concerning appropriate punishments for motoring offences in the UK.

D&L begin their analysis by taking the following two similar blends: **There is a difference whether a person drives...or somebody parks...** and **To me there is a difference whether somebody steals...DM 50.- or (he) goes to a big department store...** . They first relate these blends to these two phrases in English: **there is a difference between + NP and + NP** and **it makes a difference whether NP + VP or NP + VP.** D&L list five possible expressions in German that may also lead to two unnatural collocation blends in the L1. They give another five possible realisations of negative L1 forms, but conclude that even such a detailed post hoc error analysis cannot fully explain the student's L2 production.

D&L go through the student essay sentence by sentence and provide a commentary as to the rhetorical function of each sentence in organising the student's argument. They also expand the student's chain of ideas so that "cognitive" incoherence can be highlighted in how each sentence relates to other parts of the essay. They argue that occurrence of collocation blends is affected not only by interlinguistic transfer but also by incoherent thinking. Looking at particular examples of blends, D&L identify two sets, intra-clausal blends and supra-clausal blends. Intra-clausal blends result from "L2 deficits" (p.156) and may or may not show L1 interference. The following examples illustrate the types of intra-clausal blends, with the competing syntagmas also shown:

(1) Prepositional phrases
Example: **in context with**
Competing syntagmas: **in comparison with / in the context of**

(2) Adjective complementation
Example: **he has got financially broke**
Competing syntagmas: **he is broke / he got into financial difficulties**
(3) Verb complementation
Example: is easy to be answered
Competing syntagmas: **is easy to answer / is easily answered**
(4) Noun phrases
Example: **the best thing to handle the problem**
Competing syntagmas: **the best thing to do / the best way to handle this problem**

In sum, seven examples of intra-clausal blend and three examples of supra-clausal blend are reported. These are taken from other essays in the corpus and result from a link to another sentence in the essay. One example is:

Example 1: In this case **the limited ability of judgement** is not a problem because more serious offences are normally discussed in court with the aid of lawyers.
Other sentence referred to: **It may be that quite a number of problems, such as *the ability of police officers to judge a motorist's income*, are still unsolved.**
Competing syntagmas: **the limited ability of police officers to judge a motorist's income / the limited judgement of police officers**

D&L attribute wider causes to the frequent occurrence of blends. First, they note that cognitive incoherence may cause blends in the L1 as well as in the L2. Second, German students of English have had greater training in L2 syntax than L2 collocations. Third, they have had very little experience of writing on complex themes in English, so they use an extremely careful style in such writing. In short, L2 collocation blends cannot be fully explained by L1 interference alone.

Commentary: I'd like to consider three points of interest: (a) the attested evidence for collocation blends, (b) the elicitation task, and (c) measures of collocation competence. The first point is that the construct of collocation blend, which is derived from general linguistic theory, cannot be measured without retrospective re-construction. The Saussurian notion of syntagma is fundamentally a concept of text production: "(it) acquires its value only because it stands in opposition to everything that precedes or follows it, or to both" (de Saussure, 1915, p.121). However, we have something different here in D&L's study, where there is no attested evidence

for the underlying competing syntagmas proposed in the production of intraclausal blends. Rather, we have D&L's sophisticated detailed reconstructions of each instance. This leaves us with something of a problem for determining whether in fact the student writer did or did not have access to any two syntagmas at the time of writing. D&L's data analysis authoritatively dissects the learners' writing, but we do not hear the learners' side of the story. Think-aloud protocols could, for example, be useful in L2 collocation research.

The second point of discussion concerns the use of the written prompt. D&L are aware of how the elicitation task may affect the quality of collocation performance by learners. In using a written prompt, they argue that a careful style is encouraged, particularly with the sophisticated essay prompt they used. Although writing essays is a common way of eliciting examples of learner language, it does somewhat disadvantage the fluent production of the L2 and thus affects the quality of the collocation performance. In a later study, Lennon uses repeated recordings of a picture-story narration to analyse advanced learners' lexical errors (Lennon, 1996), suggesting that oral elicitation tasks emphasizing vernacular style can be potentially useful.

The final point of discussion is the measure of collocation. We can readily see that the term collocation blend covers many syntactic structures that may well be idiosyncratic. We do not know if the examples are representative of other similar examples or simply mirrors unto themselves. This is a weak point in the study. If we cannot generalize through observing repeated occurrences of the same describable phenomenon, then we are no nearer to a model of L2 collocation acquisition and knowledge (cf. Stubbs, 2001).

Conclusion: This small-scale study takes a narrow look at collocation blends and presents an in-depth post-hoc analysis of a handful of items. The evidence is limited, and the study claims that mis-collocations cannot be explained by L1 interference alone. Rather, it seems that learners may sometimes combine two competing L2 syntagmas to create such blends.

2.3.2 Study 8: Chi, Wong & Wong (1994)

Summary: This study used a one million word sub-corpus from the HKUST Learner Corpus, consisting of approximately two thousand 500-word essays written by first-year undergraduates. The analysis was centred on collocation errors in V + N collocations with the delexical verbs **have, make, take, do**, and **get**. Through concordance analysis, the targeted combinations were extracted, cross-checked with three dictionaries, and then checked by several NS to ensure that the final list of erroneous collocations was valid. In total, 167 combinations were identified (12 combinations with **have**, 16 with **do**, 44 with **make**, 46 with **take**, and 49 with **get**). The erroneous collocations were then put into two groups. The first group covered errors which came from wrong choice of one of the 5 delexicalised verbs. The second group consisted of learners using a delexicalised verb instead of another lexical verb.

Results from the first group showed that **do** was most confused with **make**, **get** with **have** and **make**, **have** with **make**, **make** with **have** and **do**, and **take** with **have**, **make** and **do**. Specific examples include ***do effort** (= make effort), ***get problem** (= have problem), ***get solution** (= have solution), ***have decision** (= make decision), ***make relationship** (= have relationship), ***make research** (= do research), ***take interview** (= have interview), and ***take research** (= do research). The second group involved such instances as ***do help** (= give help), ***get goal** (= achieve, reach, attain goal), ***have language** (= speak language), ***make emphasis** (= place, put, lay emphasis), and ***take challenge** (= take up, accept).

Four possible causes for the mis-collocations are proposed. With a limited awareness of the nature of delexicalised verbs, students will "choose the wrong verb-noun collocation unless they have previously learned it as a chunk" (p.162). The second possible cause concerns L1 influence. Here CW&W give several examples of how the mis-collocations involve direct Chinese L1 verbal equivalences in the L2. The L1 verb is literally translated as **do** for such

nouns as **effort, exchange**, and **results**. The third cause is related to phonological and lexical simplification, where the authors speculate, for example, that **'ve got** may have been reduced to **get** in ***get confidence** and ***get the right reasons**. The final cause is traced to reduction of a phrasal verb to a single verb. This is shown with examples such as ***take challenge** (= take up challenge) and ***take the post** (= take up the post).

CW&W conclude that many of the noun phrases in the mis-collocations are "rather abstract in a sense that they do not refer to any physical objects or physiological behaviour" (p.164). This makes it difficult for students to guess or generalise about suitable verbal combinations with such nouns. According to CW&W, students need to be introduced to such V + N collocations in appropriate contexts so they can be learned as chunks.

Commentary: I would like first to highlight some interesting general differences with the two studies in the previous group, which pursued a large-scale collocation inventory line of investigation. I will then narrow my comments down to the question of mis-collocations.

The first difference with the large-scale analyses conducted by Zhang and Gitsaki is that CW&W work from a large corpus but confine their focus to 5 delexical verbs. This allows them to retrieve recurrent examples of clear delexical verb collocations, so their analysis has greater face validity. Secondly, although they formally acknowledge Benson's division of lexical and grammatical collocations, they concentrate on one particular pattern of a single lexical collocation type, namely delexicalized V + N collocations. This leads to a more coherent analysis that does not run into some of the methodological problems found in Zhang and Gitsaki's wider ranging experiments. The other striking difference is that CW&W look at mis-collocations only, which, in some respects, gives their study a rather one-sided emphasis.

The exclusive focus on mis-collocations focuses attention on deviant forms as the way into understanding learners' collocation knowledge. If we were also to look at instances of well-produced delexicalised V + N collocations, we would have a contrastive set of data for

identifying similarities and differences between error and performance. This might not only affect how we understood mis-collocations, but also reduce the tendency to generalise from errors that we can observe. An example of unfounded generalisation is the authors' contention that learners will "choose the wrong verb-noun collocation unless they have previously learned it as a chunk" (p.162). We have no direct evidence from the data in this experiment that this holds true, so we have an unverifiable cause attributed to an effect in the data. Do learners at whatever level of proficiency learn collocations as chunks? Do they store them as chunks or unanalysed parts, and then retrieve them in the same form as stored? These are some further questions that this study raises. The same line of thinking could also be applied to one of their other findings, too. CW&W suggest that some of the mis-collocations involve direct equivalences between Chinese and English. Again, it would be interesting to see correctly produced delexicalised V + N collocations and to find out if they did or did not include direct equivalences. If they did, then we would be able to question more closely the interlinguistic assertion made about the mis-collocations found in this study.

The final point of interest is that the authors notice how some of the mis-collocations involve nouns that do not directly refer to "any physical objects or physiological behaviour" (p.164). We have an echo here of Dechert and Lennon's point about "conventional collocations" being arbitrarily fashioned by particular language groups. A further implication is that L2 learners run into collocation problems because they have a poorly developed sense of collocation restriction.

Conclusion: This study analyses 167 delexical V + N mis-collocations from a large essay corpus. The results show that learners tend to confuse delexical verbs in the L2, and various causes for such mis-collocations are proposed. However, the study does not look at correctly produced delexical V + N collocations, so we lack baseline data for making sense of the results.

2.3.3 Study 9: Granger (1998)

Summary: The NNS data for Granger's study come from the International Corpus of Learner English (ICLE) project (Granger, 1998a), which is comprised of sub-corpora of writing by English L2 learners with different L1s. The NNS corpus consisted of 251,318 words written by French L1 speakers. Granger also used a NS corpus of 234,514 words comprised of essays from 3 native speaker corpora. Collocations and formulae form the research focus: by collocation, Granger means lexical collocations (Benson, Benson & Ilson, 1986b) or restricted collocations (Aisenstadt, 1979), whereas formulae correspond to the same concept as lexical phrases (Nattinger & DeCarrico, 1992), realized in such phrases as **it seems (to me) that** and **be that as it may**.

Granger extracted **–ly** amplifiers and formulaic sentence builders for Contrastive Interlanguage Analysis. First, raw frequencies of such amplifiers were examined. The NS type/token raw frequencies total was 75/313, while the NNS was 41/230. Results showed that **completely** and **totally** tended to be used much more frequently by the NNS group than **highly**, while the converse was true for the NS group. Granger attributes this overuse to the French speakers using **completely** and **totally** as "safe bets" (p.148); moreover, these two amplifiers have direct L1 equivalents which can be widely used without collocation restriction. **Hautement**, however, has a much more restricted register and is less frequently used in general. The second step was to examine the use of maximizer amplifiers like **absolutely** and **totally**, as well as boosters such as **deeply** and **strongly**. Boosters were notably underused by the NNS writers. Granger found that the NS exclusive boosters could be classified as either stereotyped combinations such as **vitally important** and **readily available** or creative combinations. Both types were underused by the NNS group, and the rare stereotyped combinations that did occur in the NNS data seemed to have direct L1 equivalents. The tendency of the NNS group to use "general-purpose" items as amplifiers was further highlighted by the extreme overuse of **very**.

Granger ran an additional test with 56 English NS and 56 French NS. The informants were asked to make appropriate matches between 15 adjectives and 11 amplifier collocates. The adjectives were: **significant, reliable, ill, different, essential, aware, miserable, available, clear, happy, difficult, ignorant, impossible, cold,** and **important**. The amplifiers consisted of: **highly, seriously, readily, blissfully, vitally, fully, perfectly, heavily, bitterly, absolutely,** and **utterly**. Each amplifier was placed at the front of the list of adjectives in bold, and for each amplifier the subjects had to read through the adjectives and select possible collocates. The same list of adjectives was repeated each time. The English NNS group marked 100 fewer possible combinations than the English NS group (280 vs. 384, or an approximate mean ratio of 5:7). They also chose a greater number of different combinations, overgeneralising the use of **highly** and **fully**.

Granger's second area of focus was contrastive analysis of the use of sentence builders, i.e., "phrases which function as macro-organizers of the text" (p.154). She found that the learners greatly overused active frame sentence builders such as **We/one/you can/cannot/may /could/might say that ...**, compared to passive frames such as **It can be claimed that ....** Granger points out that unnative-like language production needs to be seen as involving "excessive use of prefabs" (p.155), not just absence of prefabs. She hypothesizes that learners stick to certain sentence builders that they can confidently use within their limited repertoires, suggesting that L2 learners resist segmentation in that they tend not to develop gestalt forms (i.e., chunks and phrases) into analytic language. Granger also comments that "the value of introspective tests in this field should not be underestimated" (p.159).

Commentary: I would like to take up three points from Granger's study: (a) the general design of her study, (b) the problem of interpreting raw frequencies, and (c) mis-collocation as evidence of over-generalisation. Granger moves beyond the one-shot design of many of the studies under review here by examining ICLE sub-corpora for the first stage in her research, analysing the

results, and then going back to two similar groups to test further the insights that she has gained. This comes across as an efficient way to validate and expand on preliminary data and re-test data-driven hypotheses. At the same time, a slight word of caution is warranted as to the differences that Granger notes between the NS and the NNS groups. For example, with the use of amplifiers, it is not clear if the NS and NNS type/token raw frequencies (75/313 or 23%, and 41/230 or 17%, respectively) constitute a reliable difference or not. We also do not know how many different writers' essays were used to establish this baseline information. Yet, in the follow-up elicitation task, Granger is able to use two groups of equal size, so the second set of results seems more solid, and the claims of over-generalised use by the NNS appear well-founded. Despite this, it is unclear how easy it is for subjects to indicate connections between 15 adjectives placed to the right of an individual amplifier collocate. To work through a list of 11 amplifiers 15 times seems a rather repetitive task of questionable validity. It is moreover unclear how the adjectives have been chosen, so once more we run into the issue of how to select appropriate items for experimental work. My final point is that Granger arrives at a view of L2 collocation use that is not deficit-based. This is because she does not solely interpret mis-collocations in negative terms. Her interpretation sees over-generalised use as evidence of "safe bets" by learners with limited L2 lexicons, so she indirectly relates collocation production to L2 vocabulary size and communicative proficiency.

Conclusion: The study establishes some working hypotheses from essay data, and a follow-up de-contextualised elicitation instrument is used to probe further. By focusing on a small set of items and administering these to two equal-sized NS and NNS groups, Granger finds that the NNS tend to over-generalise their use of both types of lexical construct. She characterizes this over-use as a "safe-bet" strategy by second language users who have limited repertoires and a restricted sense of collocation salience.

2.3.4 Study 10: Howarth (1998)

Summary: The study focuses on lexical V + N collocations in a combined NS corpus of 180,000 words of social science academic writing, and a 25,000 NNS corpus of individual postgraduate papers by 10 MA students with various L1s. The NS data were tagged for verb lexemes, and a total of 63 verb lexeme types were identified from a total of 5,379 tokens. These were then sorted into free combinations, restricted collocations, and idioms. With the NNS data, each paper was separately analysed for V + N collocations. These were lemmatized and coded with the same patterns as the NS data, before being checked for deviant forms. This produced 1,165 tokens for further analysis. Howarth refines the category of collocation restriction by considering the degree of specialization of the verb and the degree of substitution possible in one or both components. This leads to a '5-level' classification of selection restrictions (see Appendix 2.2), where Levels 2-5 are sub-divisions of restricted collocations. These are further sub-classified by semantic specialization into figurative, delexical, and technical senses.

Results showed that restricted collocations and idioms covered about a third of the total lexical collocations in the NS data. A small number of deviant forms (30) was also found. With lexical mis-collocations, Howarth identified two main types, collocation overlaps and collocation blends. Collocation overlaps occur at Level 2 where the noun objects "have partially shared collocability" (p.174). Thus, in the case of ***The contrast is drawn again a few pages later...**, it is claimed that the mis-collocation ***draw a contrast** comes from within the following overlapping set: **draw a distinction**, **make a contrast**, and **make a distinction**. Blends differ from overlaps in that they involve "two distinct collocations with no lexical overlap" (p.175), as with ***appropriate policy to be taken with regard to such inspections**. Here, **take steps** and **adopt a policy** are blended together to produce two distinct mis-collocations, ***adopt a step** or ***take a policy**. Blends may suggest that "semantically similar lexical complexes are stored nearby in the mental lexicon" (p.176).

NNS individual variation in style was stronger. Relatively more Level 2 restricted collocations were produced (63%) compared to the NS data (41%), and correspondingly fewer Level 3 combinations were found (21% versus 37%). Most lexical collocation problems could be traced to the question of semantic specialization of the verb and restrictions on the verb in combination with particular nouns. Delexical verbs came second here to figurative verb use, with restriction problems of figurative verbs related to collocational overlap and blends. Examples included ***make a reaction**, where **give a reaction** and **make a response / make a comment** may be overlapping, and ***suffer difficulty** (**suffer pain** and **experience difficulty / pain**). Blends were fewer. One instance is ***achieve tasks**, which is seen as a blend of **achieve goals** and **perform tasks**.

Howarth concludes that blends indicate competence in that writers "can only produce blends of collocations that are known" (p.185). NNS writers are generally competent with free combinations and idioms, but do not achieve full native-like phraseological competence because they have weak understanding of collocation restriction.

Commentary: I will discuss two issues: (a) the problem of corpus size for L2 data, and (b) collocation overlaps/blends as an indication of collocation competence. Howarth's L2 data are drawn from a 25,000-word corpus of writing by international postgraduates. Through applying the 5-level categorisation to the lexical V + N collocations in the corpus, Howarth ends up with a very limited number of tokens for analysis (see Table 2.3 below).

Table 2.3 NNS lexical collocations results in Howarth (1998)

Type	Token	Mean per student
Free collocations	857	85.7
Restricted collocations level 2	185	18.5
Restricted collocations level 3	63	6.3
Restricted collocations level 4	26	2.6
Restricted collocations level 5	22	2.2

We should note that evidence of lexical substitution is found in only 6% of the total NNS collocations produced, "or approximately eighty forms" (p.180). This averages out at eight per writer. It is clearly very difficult to get enough examples of lexical collocations through written tasks, so the results have limited generalizability. Separating restricted collocations into four levels with such a limited amount of data may therefore be using too fine a net to filter the data. It might be more useful to look more closely at all correctly produced V + N lexical combinations—a point that was raised in the discussion of the earlier Hong Kong study.

The second area of discussion is Howarth's assertion that the occurrence of blends may indicate collocation competence in that a person can only produce blends if they have knowledge of the two competing collocations. With **give a reaction** and **make a response**, Howarth suggests that, because the two nouns are semantically similar, "...there is a short diversion from the normal process of selecting a collocate, caused by the proximity of the two nouns in the mental lexicon" (p.182). We could also argue that **have** and **get** are also competing within the collocation set, and that **make** is combined to overcome competing signals, with or without reference to **give**. That is, the nouns are semantically similar to each other and share a small set of potential delexical verbal combinations. Howarth's representation of the blend/overlap suggests a very limited network of lexical items with activated connections, but an alternative view could claim greater competition between other potential collocates.

If we consider the example of ***suffer difficulty** (**suffer pain** and **experience difficulty / pain**), we might further claim that the collocation selection involves choice of a lexical verb **suffer** rather than delexical **have**. Howarth states that the two nouns are "considered to be semantically too dissimilar to be easily confused in the lexicon, and there exists a verb which collocates with two of the nouns in the hypothetical set" (p.182), but it is clear that the delexical verb **have** is also a common collocate for both nouns (**have difficulty/have pain**). A different interpretation might then suggest L2 users' preference for

full lexical forms, as well as a resistance to using delexicalized items.

Conclusion: The study involves a highly detailed post-hoc reconstruction of how NNS produce deviant restricted mis-collocations. The number of instances of mis-collocation investigated is limited, and the explanations proposed for how the L2 mental lexicon works are somewhat tenuous.

2.3.5 Study 11: Nesselhauf (2003)

Summary: Nesselhauf's opening argument is that the operationalization of collocation for restricted V + N combinations needs to be simple. She uses two criteria by which to decide if V + N collocations are restricted:

> **Criterion 1** The sense of the verb (noun) is so specific that it only allows its combination with a small set of nouns (verbs).
>
> **Criterion 2** The verb (noun) cannot be used in this sense with all nouns (verbs) that are syntactically and semantically possible. (p.225)

If one or both of these criteria hold(s) true, then the collocation can be regarded as restricted. In Criterion 1, "small set" means up to five elements (p.227). For example, since **read** can go together in its primary sense with "all nouns denoting written or printed matter" (p.226), **read a newspaper** can be considered a free combination, whereas the verb **dial** has a very limited number of nouns, as in **dial a number.** Three major categories of collocation—free combinations, collocations where the sense of the verb is restricted, and idioms—are proposed (see Appendix 2.3 for details). In this three-way 'restricted sense' categorisation, the use of a restricted verb with an unrestricted noun distinguishes collocations from the other two categories. Nesselhauf cautions, though, that the system of classification is not problem-free.

Thirty-two argumentative essays were selected from the ICLE corpus (Granger, 1993). Each essay was about 500 words long on average. First, each V + N combination was extracted and coded **F** (free), **R** (restricted), or **I** (idiom). The 1072 combinations were then evaluated for their acceptability by reference to two dictionaries (*Oxford Advanced Learner's Dictionary*,

2000, and *Collins COBUILD English Dictionary*, 1995). If the verbs did not indicate restriction of the verb, a combination was coded **F**. Where there were restrictions shown, a combination was coded **R**. If the dictionary indications were unclear, a combination was coded **RC?**. If further cross-checking with the BNC corpus and two native speakers showed the combination to be clearly acceptable, it was then re-coded **RC**. Those combinations with incorrect verb choice were also given a NS check, with the correct verb supplied by the NS.

The re-constituted combinations were then evaluated as to whether they were **F**, **R**, or **I**. Acceptability judgments were first made against information in the dictionaries and the BNC. Where it was still not possible to reach a decision, combinations were checked by two NS, who rated the combination as **C** (correct), **W** (wrong) or **CW** (not sure). If there was still no agreement, the combinations were run past a further two NS. This led to a 5-degree cline of acceptability: clearly acceptable = **C** / largely acceptable = **(C)** / unclear = **CW** / largely unacceptable = **(W)** / clearly unacceptable = **W**.

Results showed that approximately 25% of all combinations (255) featured errors, of which 56 were restricted combinations. In these 56 mis-collocations, three main types of error could be observed: 24 involved the wrong choice of verb, 14 the wrong choice of noun, while 11 involved combinations that don't exist conventionally. These **RC** combinations were then re-coded into **RC1** (involving a lot of restriction) and **RC2** (involving only a little restriction) types. Twenty-nine of the mis-collocations were **RC2** types, and 14 **RC1**s. According to Nesselhauf, "whereas learners are mostly aware of the restriction in combinations where the verb takes only a few nouns, they are less aware of restrictions where the verb takes a wider range of nouns" (p.233). Through further analysis of the lexical congruence between the L2 combinations and their L1 versions in German, on average 45% of the mis-collocations seem to have been subject to L1 influence. For correctly produced combinations, Nesselhauf also checked lexical congruence between the L1 and the L2 and found that the more restricted combinations were

less lexically congruent. Noting that L1 influence appears to be stronger than previously assumed, the paper closes with pedagogic suggestions for dealing with collocations.

Commentary: I would like to comment briefly on the following three points in Nesselhauf's study: (a) the limited generalizability of the study, (b) the differences in verb and noun knowledge, and (c) the emphasis on lexical congruence.

The first point concerns the tension between conceptual simplicity and reliability. Working from the same notion of a collocation continuum (**free/restricted/idiomatic**) as Howarth, Nesselhauf simplifies the middle area of restriction to whether the verb is used in a restricted sense with particular nouns. This follows through from Howarth's finding that most mis-collocations could be traced to the use of the verb. Having simplified the focus for analysing her data, she is able to retrieve an apparently large number of restricted collocations. Yet, of the 255 examples, just 56 involve restricted mis-collocations. So, although the corpus consists of 32 essays totalling 16,000 words, there are, on average, fewer than two examples per writer of restricted mis-collocation. This once more underlines how difficult it is to obtain from corpus data sufficient examples to achieve reliable, generalizable results.

Secondly, Nesselhauf's results tend to indicate that noun knowledge is stronger than verb knowledge, which supports the theoretical premise that the dependent part of a V + N collocation is the verb. As this is based on a very limited number of examples, it deserves further investigation. Nesselhauf makes a number of inferences from these results about learners' collocation awareness; it would be similarly useful to test these claims by conducting structured interviews with learners so that finer insights could be developed.

The final point that I would like to discuss is Nesselhauf's conclusion that 45% of the 56 mis-collocations can be explained by L1 influence and lack of lexical congruence between German and English. While this may be the case, it does not explain the reasons for the other 55%. This is a curious oversight in a tightly organised and theoretically well-grounded paper.

Conclusion: This study investigates restricted collocations in a small corpus of advanced written English. It provides a detailed account of what collocation restriction is and how collocation restrictions operate in V + N combinations. The results indicate that learners have more problems with selecting the appropriate verb than noun in such collocations. However, the number of examples identified per subject is highly limited.

2.3.6 Concluding comments on the essay corpus with narrow analysis of specific collocation types studies

The data in the five studies in this group range from the very small scale of the Dechert and Lennon study (seven intra-clausal blend example and three examples of supra-clausal blends within one student essay) to the one million-word corpus of the Hong Kong study. All five studies indicate the difficulty of obtaining enough instances of the particular collocation construct under investigation. Granger's study is the only one to address this problem by moving from corpus data to a second stage of de-contextualised experimentation. So, the first characteristic of these studies is the limited generalizability that the results from corpus analysis alone provide.

The second common feature of these studies is that they tend to limit their focus to particular types of lexical collocation, most notably V + N combinations. The studies tend to be conceptually clearer than the investigations presented in the previous two groups. The finer focus also results in a more detailed analysis of particular types of collocation, although a great deal of the analysis tends to involve secondary reconstruction of how learners produce collocations and what interferes with their native-like production. Here, the studies tend to diverge in their interpretations. Mis-collocations are seen as resulting from either a lack of L1-L2 lexical congruence or over-generalisation in the L2. The first interpretation tends to come from limiting the analysis to mis-collocations alone, while the second interpretation takes account of both correctly and incorrectly produced collocations.

Although this difference may seem slight, the implications affect the way in which the learner is positioned in such L2 research. Is the learner to be seen as merely the disembodied producer of interesting but deviant language, or is there a further role for learners to play in shedding light on how they use language? Granger is the clearest about this when she concludes that informant introspection should not be underestimated as means of better understanding collocation use and development. Granger's study is also the only one where a second experimental stage is included.

2.4 Experimental measures of collocation knowledge

The final group involves four studies aimed at quantifying collocation knowledge through experimental measures (Read, 1993, 1998; Schmitt, 1999; Mochizuki, 2002). Read (1993) initially focuses on three types of lexical knowledge—syntagmatic, paradigmatic and analytical—and then narrows down to syntagmatic and paradigmatic in a revised version of his word associates test (Read, 1998). Schmitt isolates four specific types of lexical knowledge and measures these concurrently, whereas Mochizuki examines both paradigmatic and collocation knowledge of 72 individual words. The last two papers correlate their results to overall measures of lexical competence, while Read uses an independent measure of academic vocabulary knowledge for correlation. The four studies in this group thus treat L2 collocation knowledge in relation to other types of word knowledge and overall lexical proficiency.

2.4.1 Studies 12 and 13: Read (1993, 1998)

Summary: As Read's 1998 paper is a validation of his 1993 study, the two studies by Read are presented here as a single summary and commentary. The starting concern is to find a test format that can provide broad coverage of depth of knowledge of academic vocabulary in a time–efficient manner. This is attempted through using word associates, i.e., items which may be considered part of the lexical set of a particular word. The initial format of the word

associates test (WAT) consisted of stimulus words with eight possible associates, of which four were distractors. An example item is:

edit

arithmetic	**film**	**pole**	**publishing**
revise	**risk**	**surface**	**text**

According to Read, one associate is paradigmatic, i.e., a near-synonym (**edit-revise**), two are syntagmatic, i.e., collocates (**edit-film** and **edit-text**), and one is analytic, i.e., part of the wider definition of the stimulus (**edit-publishing**). Moreover, the four distractors (**arithmetic**, **pole**, **risk** and **surface**) have been deliberately chosen because they have no semantic association with the stimulus. Two 50-item forms were produced, with about 30% of the associates taken from the UWL (Nation, 1990). The two forms were piloted and revised, before being administered to international students at Victoria University (n=103 for Form A; n=112 for Form B).

Results showed that the WAT had good reliability (KR-20: 0.92) and that the two forms correlated strongly with each other. Item Response Theory analysis identified three misfitting items on Form A and Form B, and 14 misfitting subjects overall. It was not possible to interview these learners to find out directly why, but Read did conduct separate think-aloud protocols with eight learners as they completed Form A. Results from this qualitative analysis indicated that learners tended to mark off those associates that they knew and look more carefully at the remainder to make an informed guess where they were unsure. If the stimulus item was unknown, higher proficiency learners were more likely to guess, while less proficient learners refrained from answering. The think-alouds also revealed that test-takers could sometimes identify word associates correctly even if they did not know the stimulus word. Finally, good correlations were also established between the WAT and a separate measure of academic vocabulary knowledge.

The later version of the WAT presented two sets of four words for each stimulus item. One example is:

sudden

beautiful quick surprising thirsty	**change doctor noise school**

The set on the left focuses on paradigmatic relationships with the stimulus word (**quick** and **surprising**), and the one on the right on collocation links (**change** and **noise**). The revised version kept the number of correct responses per stimulus fixed at four per item, but varied their distribution between the two sets. A particular set might therefore contain one, two or three correct associates. This feature was introduced to cope with the guessing behaviour that the earlier think-aloud analysis had shown; it was also necessitated by the varying number of plausible collocations that the selected stimulus adjectives had. In the revised version, all stimulus items and syntagmatic associates were adjectives, and the collocation links were restricted to nouns. The adjectives were mostly taken from the 2000 and 3000 word lists of frequent vocabulary rather than just the UWL.

One 46-item form was prepared and piloted, and a final revised 40-item instrument was used for further validation. Two concurrent measures were run, individual interviews and an enlarged matching test (MT) of 80 items. In the MT, each of the 40 WAT stimuli was tested twice for two different basic senses. In the first trial, 84 subjects took the MT and the WAT; in the second, 38 subjects took both tests, and 15 learners (12 'higher proficiency', three 'lower proficiency') were interviewed about their knowledge of 20 stimulus adjectives from the WAT. Read used a modified version of Wesche and Paribakht's Vocabulary Knowledge Scale (1996) to score the interviewee responses (see Appendix 2.4).

Results from Trial One showed that both tests had good reliability (.93 for the WAT and .90 for the MT) and that the two forms of the test correlated strongly with each other (.82). Rasch analysis identified one misfitting item in the WAT and four misfitting subjects overall (who had left answers blank as they seem to have been reluctant to guess). The findings from Trial Two produced large SDs for both tests. This was attributed to the wide range of proficiency and small population. The two tests had a high correlation co-efficient of .86, and

the MT had a higher correlation than the WAT with the interview results (.92 vs. .76). Read analysed the consistency of individual responses for 20 adjectives across the three measures. In 69 out of 300 cases, the criterion level was reached on the WAT only, but not on the other two measures.

Commentary: Read's work is different from the other studies reviewed in this chapter in that it focuses on designing, running and validating a particular measure of vocabulary knowledge through both quantitative and qualitative analysis. The instrument is based on a 'depth of knowledge' interpretation of lexical knowledge and tests a relatively large number of items in order to estimate the overall academic lexical proficiency of international university students. Two important issues arise here for discussion: the construct of depth of lexical knowledge and the lack of overall response consistency at the end of the validation process.

Read puts forward a two-part definition of depth of knowledge as "the extent to which learners are familiar with the range of meaning and uses of a set of target words" (Read, 1998, p.41). Although it is not absolutely clear, I take the revised WAT to be simultaneously measuring range of meaning (the left-hand set of adjectives) and word use, i.e., collocation (the right-hand set of nouns). As the WAT has these two separately operationalized parts, it follows that we should expect two separate scores—one for paradigmatic knowledge, and the other for syntagmatic knowledge. However, that is not the case: there is a single score presented for the WAT results, which suggests a unitary view of depth of knowledge. In other words, although different aspects of the construct of depth of knowledge are advanced, the instrument measures them as a single dimension. One reason for this may be practical. As noted above, the revised WAT had a varying distribution of correct associates for the two sets of responses for each stimulus adjective; it might simply be that the distribution was unequal in the test overall, so that there are two different totals for correct paradigmatic and syntagmatic responses. If this is the case, then the revised WAT has a major weakness in its internal validity. A second effect is that

we cannot in any case derive a clear score for L2 collocation knowledge, so the relevance of the instrument to measuring L2 collocation knowledge remains obscured.

Read acknowledges the problem in a later discussion of the necessary trade-offs inherent in designing any measure of vocabulary (Read, 2000, pp.178-187). However, the interview data from Trial Two of the revised WAT suggest that the guessing behaviour of test-takers remains strong. Associative knowledge (or depth of knowledge) is not very neat and tidy to measure despite the careful revisions made during the validation process. In short, the WAT is premised on the notion of partial depth of knowledge of words, but the final version of the WAT falls some way short of accounting for this in the way that it is designed and scored.

Conclusion: The two studies by Read are exceptional in the attention shown to the design, administration, and validation of a measure of depth of lexical knowledge. The research exploits both quantitative and qualitative analyses to investigate how the instruments work. The final version of the WAT achieves very good results, but the analysis of response consistency raises major questions about what those results in fact show.

2.4.2 Study 14: Schmitt (1999)

Summary: Schmitt notes several problems with the way the TOEFL test tries to measure vocabulary knowledge. First, TOEFL guidelines do not specify what kind of lexical knowledge is tested. Second, in trying to estimate the construct validity of TOEFL vocabulary items, the predictive validity of such items is weak because they are subsumed within a test of reading ability and cannot easily be correlated to future vocabulary learning. Third, other tests such as the Vocabulary Levels Test and Yes/No tests, which focus on vocabulary size, have yet to be validated and compared with how the TOEFL vocabulary test items work. This means that the construct validity of the TOEFL test is weak, and its predictive and concurrent validity unclear. To establish a better sense of the construct validity of the TOEFL vocabulary items, Schmitt refers to Nation's typology (Nation, 1990; 2001, p.347) for different types of word knowledge

(the word and its spoken form, written form, grammatical behaviour, collocation behaviour, frequency, stylistic constraints, conceptual meaning, and its associations with other words). As measuring all those different types of word knowledge would be impractical, the study limits its focus to meaning, word association, collocation, and grammatical word class.

Six target items "with the greatest number of different meaning senses" (p.196) were selected. These were: **massive**, **peak**, **rare**, **subtle**, **surging**, and **trend**, which were found to be used in their most typical sense on the TOEFL (with the exception of **massive**). The items were normed for the four types of word knowledge. Meaning and word form were checked against three dictionaries; word associations were obtained from 50 NS, and the COBUILD Bank of English was consulted to establish three target semantic fields with their respective collocates for testing collocation knowledge. The items were then run past 30 pre-sessional international students in the following sequence in individual interviews: (a) word association test, (b) TOEFL test, (c) word class, (d) prompted sentence-level composition to elicit collocations, and (e) meaning senses. To elicit collocations, the subjects were required to produce orally sentences with the target items. The following example shows how the collocation prompts were used:

PEAK
1. **If you were talking about a business**
2. **If you were talking about a house**
3. **If you were talking about geography**

Although the target collocates are not given in this 1999 paper, another paper by Schmitt (1998a) lists the normed collocates extracted from the Collins-COBUILD Bank of English. For the three different senses of **peak**, the target collocates were:

PEAK
Business: ***career, demand, levels, season***
House: ***roof***
Geography: ***Himalayan, mountain, snow, top***

Results showed that answering a vocabulary item correctly on the TOEFL did not necessarily mean that all four types of word knowledge had also been fully acquired. Subjects generally knew the most common sense of a word, but had weak knowledge of other meaning

senses. Even when the learners could not show that they knew the meaning of a word at all, they could in some cases still answer the TOEFL items correctly. As for word association knowledge, when TOEFL items were accurately answered, in just a third of cases could the subjects produce native-like associations. They did better on word-class form, where they could identify the correct word class in 87% of cases and supply two or more derivatives in 69% of instances. Subjects nevertheless showed much weaker collocation knowledge of the target items. In 36% of cases, they failed to compose any sentences with the target collocates, although most managed to supply at least one example of the collocation range of the six words. In sum, word form knowledge was found to be stronger than word association, sense, or collocation knowledge.

Commentary: I will limit my comments to how collocation knowledge is elicited. The collocation elicitation task in this study makes several assumptions which are worth unpacking. The design of the collocation prompt implies that collocation relationships are activated by a topic prompt, and that a language user is mentally versatile enough to be able to produce at least one of the target collocates for each of the three topic prompts. The assumption is that NS can do this across all three topic prompts and demonstrate 'collocation normalcy' in their responses. Non-native speakers, however, will most probably show variation but will nevertheless be compared to the assumed 'ideal' NS norm. We would, in other words, expect NS to score three each time for each of the prompt words, and NNS to vary in their scores. To test this assumption in an albeit very limited way, I ran the prompts for **peak** and **trend** past a British native-speaker colleague. He scored one out of three for both sets of prompts, and showed no greater collocation knowledge than the NNS in Schmitt's study (see Appendix 2.5).

If it is the case that NS are limited in the collocates that they can produce, this raises a number of interesting questions. It might be that NS have conscious access in their mental lexicons to one or two primary collocation relationships only, and that these are highly

conditioned by recent exposure or use. In other words, NS may find it quite difficult to search for, with a topic prompt, and retrieve more than one or two expected collocates. Alternatively, it is possible that the prompts used in this study are not strong enough in triggering expected collocates, despite the corpus evidence suggesting that there is a set of normal collocates which co-occur with the target items. That is one major weakness of the test of collocation knowledge: the norms are drawn from a massive amount of *written* corpus data, but, unlike the word associations in this study, have not been tested against a group of NS in similar conditions of *spoken* elicitation to the non-native speakers. This may inadvertently skew the results and obscure the possibility that the NNS group may have been doing better than judged.

Conclusion: This exploratory study attempts a concurrent measure of four specific types of word knowledge and uses innovative procedures to measure L2 collocation knowledge. Although collocation knowledge is found to be the weakest of the four types of word knowledge, the experiment suffers from the very limited number of items used and from partial NS piloting.

2.4.3 Study 15: Mochizuki (2002)

Summary: Mochizuki's study aimed to investigate over a period of nine months possible correlations between vocabulary size and paradigmatic and collocation aspects of lexical knowledge. The experiment involved 54 first-year students majoring in German, Chinese or Japanese, who had two 90-minute classes of English a week for 25 weeks. They had previously learned English for six years. The test instruments consisted of a modified version of Nation's VLT test (with the definitions given in Japanese, and divided into seven 1000-word bands); a 72-item paradigmatic test of vocabulary knowledge (grouped into four sets of 6 nouns, verbs and adjectives, arranged by frequency); and, a collocation test targeting the same 72 lexical items as the paradigmatic test. For the vocabulary size test, the subjects had to choose the appropriate Japanese definition from sets of six definitions for each group of three words. The paradigmatic test required the subjects to select the most appropriate synonym, superordinate, co-ordinate, or

subordinate from a set of four for each target item. The following examples are given:

job	**(1) date**	**(2) sort**	**(3) star**	**(4) work**
claim	**(1) affirm**	**(2) expire**	**(3) mimic**	**(4) transit**
available	**(1) accessible**	**(2) chronic**	**(3) latent**	**(4) notorious.**

In the collocation test, the goal was to select the appropriate collocate from a group of four possible choices. Where the target item was a noun, verb or adjectival collocates were given; where the target item was an adjective or verb, the choice was restricted to noun collocates, e.g.,

job	**(1) answer**	**(2) find**	**(3) lay**	**(4) put**
claim	**(1) contrast**	**(2) generation**	**(3) responsibility**	**(4) sorrow**
available	**(1) debt**	**(2) economy**	**(3) information**	**(4) surface.**

The collocates were taken from COBUILD and the Edinburgh Associative Thesaurus. The vocabulary size test was administered in the first class of the year; a week later, the two other tests were given, before all three tests were completed again at the end of the academic year.

Results showed good reliability for the paradigmatic test on both interventions (α 0.71 in April, and α 0.75 in January), and moderate reliability for the collocation test (α 0.54 in April, and α 0.70 in January). The average vocabulary size did not noticeably grow in the same period (from 4,848 words to 4,859 words). Scores remained stable for the paradigmatic test (mean 46.57 in April, mean 47.04 in January), and for the collocation test too (mean 41.72 in April, mean 42.81 in January). An F-test confirmed significant difference in variance ($p<0.05$) for the collocation test only. A t-test for the vocabulary size and paradigmatic tests, together with a non-parametric test for the collocation test, indicated no significant difference between April and January for the three test means. Mochizuki attributes lack of motivation and the learning of another language in the students' major as possible contributing factors in the stability of vocabulary size and paradigmatic knowledge. Following Schmitt's findings that "knowledge of meaning sense has a certain amount of inertia and does not change easily" (Schmitt, 1998, p.300), Mochizuki similarly claims that collocation knowledge may be more subject to losses or gains than word meaning knowledge.

Commentary: There are two issues that I would like to raise. The first is establishing a measure of vocabulary size. The second concerns the measure of collocation knowledge. The

modified version of the VLT that the author used contained 30 items in 15 sets for each of the seven levels (Mochizuki, 1998, pp.52-53). In each set, the learners were required to match two Japanese translations to six possible English choices. The third level of the test (VST-3) contained eight sets of nouns, three sets of verbs, three sets of adjectives, and one set of adverbials. We may note that the distribution of word classes is slightly different from Nation's original version (Nation, 1990), and that Mochizuki's modified version plays to Japanese students' strengths, particularly at the beginning of their first year at university. They will have crammed vocabulary equivalences for the university entrance exams mostly held in February in Japan, and their starting vocabulary size in terms of recognition of L1-L2 equivalents will be at its peak at the start of the academic year in April. Many students report that their English vocabulary drops in the first semester at university, as their temporarily acquired knowledge of individual items falls away, post-entrance exam. This may explain why the test of vocabulary size did not show any significant gains. More importantly, we should question the representativeness of such sampling of vocabulary size. If there are 30 items per 1,000 words, each item has a coverage of about 33 items. This is slightly better than the 5K Swansea checklist test (100 real words, 50 non-words), where each real word covers 50 words. Both tests have relatively wide representative margins, and it may simply be the case that such measures underestimate vocabulary size in the higher ranges of their sampling.

The second discussion point concerns the changes in collocation knowledge. The analysis in this study moves away from examining collocation competence by cloze, translation or written essays, which, as we have seen, tend to be the main research tools used in most other studies. The present study also seeks to measure a large number of items in a relatively short period of time. This is a further positive feature of the instrument. One part missing in the reported findings, however, is whether the changes in collocation knowledge were similar or not across the three word classes used in the prompts (nouns, verbs, adjectives). It would also be

interesting to test the correct collocations as full two-part items, such as **find job, claim responsibility** and **available information**, to see if Japanese learners claim recognition of a collocation phrase to the same degree that they can choose appropriate collocates. Mochizuki's multiple-choice approach emphasizes analytical ability rather than strength of learner recognition of collocations. It would also be useful to know to what extent the collocations targeted in the study show direct L2-L1 equivalence so that we might judge whether recognition was influenced by lexical congruence between the two languages or not.

Conclusion: This test/re-test study limits itself to correlating different types of lexical knowledge by measuring overall vocabulary size, paradigmatic vocabulary knowledge, and collocation knowledge. A large number of items is tested, but no significant differences are found in the three measures between the beginning of the academic year and the end. The tight lexical focus stands in marked contrast to many other L2 collocation studies.

2.4.4 Concluding comments on the experimental measures of collocation knowledge studies

The four studies are clearly positioned in the field of second language vocabulary studies: they treat collocation knowledge in relation to other types of word knowledge and overall lexical proficiency. An important difference between the studies is that Schmitt tests only a small number of items, because of the time constraints involved in using oral interviews with 30 subjects. However, the four papers are similar in that they use decontextualized instruments to examine collocation knowledge. In Read's case, this is done through a stimulus item with two separate sets of word associates, while Schmitt employs topic prompts and Mochizuki's test requires matching one of four possible collocates to a written prompt. Finally, although the problems with the design of the WAT do not allow Read to establish conclusive differences between paradigmatic and syntagmatic word knowledge, both Schmitt and Mochizuki identify collocation knowledge as weaker than other types of word knowledge. Mochizuki further suggests that L2 collocation knowledge is more likely to decrease rather than increase over one

academic year.

This brings us to the end of the presentation of previous L2 collocation research. To draw the four groups of studies together, it may be helpful to take a look at the key features of each set of investigations. Table 2.4 on the next page presents a bird's eye view of the 15 studies in the order that we have reviewed them. In the next section, we will discuss certain points of interest that previous research into L2 collocation knowledge raises.

Table 2.4 Key-feature summary of previous research into L2 collocation knowledge

Previous study by group	Context and L1	N	Level*	Number & type of target collocations	Measure
1. Biskup, 1992	Poland: Polish & German	34 + 28	A	23 V+N, Adj+N	Translation into English
2. Bahns & Eldaw, 1993	Germany: German	58	A	15 V+N	Translation into English, cloze, V size
3. Farghal & Obiedat, 1995	Jordan: Arabic	34 + 23	A	1 N+N, 10 Adj+N pairs	Translation into English, cloze
4. Bonk, 2000	USA: Mixed	98	PI, I	17 V+N, 17 V+Prep, 16 Fig. V	Cloze, acceptability judgement
5. Zhang, 1993	USA: Mixed	30 NS 30 NNS	I, A	66 lexical & grammatical types	Essay, cloze, acceptability judgement
6. Gitsaki, 1996	Greece: Greek	275	E, PI , I	37 lexical & grammatical types	Essay, translation into English, cloze
7. Dechert & Lennon, 1989	Germany: German	9	A	Intra- & inter-clausal blends	Essay corpus: error analysis
8. Chi Man-lai, Wong Pui-yiu, & Wong Chau-ping, 1994	Hong Kong: Cantonese	200 approx.	I	167 tokens of 5 delexical V+N types	Essay corpus: error analysis
9. Granger, 1998c	France: French	? NS ? NNS	A	230 tokens of 41 -ly amplifier types	Essay corpus: error analysis, matching
10. Howarth, 1998	UK: Mixed	? NS 10 NNS	A	1165 V+N tokens	Essay corpus: error analysis
11. Nesselhauf, 2003	Switzerland: German	32	A	1072 V+N tokens	Essay corpus: error analysis & correctly produced collocations
12. Read, 1993	New Zealand: Mixed	101 (both forms)	I, A	50 pairs of collocates	Word associates format + matching test + think-alouds
13. Read, 1998	New Zealand: Mixed	84 + 38	PI, A	?? Adj + N pairs	Revised word associates format + matching test + interviews
14. Schmitt, 1999	UK: Mixed	30	I, A	6 Adj or Ns	Prompted written elicitation of collocates
15. Mochizuki, 2002	Japan: Japanese	54	PI, I	V/Adj collocates for N, N collocates for Adj/V	Collocate selection test

* A= Advanced, I=Intermediate, PI=Pre-intermediate, E=Post-beginner/Elementary

2.5 Discussion

At the start of this chapter, we noted the weak sense of interconnection between the different studies. By grouping them into four methodologically similar sets, we were however able to identify some common points. One of the striking points about previous investigations is how they have tended to limit their focus to particular collocation types. In fact, V + N combinations have proved to be the most frequent area of concern, but no research has been conducted into how well the individual verbs and nouns of particular V + N collocations are known. A second related point is that most of the previous research has concentrated on examining collocation production. Few studies have tackled the question of collocation recognition. A third issue is the lack of agreement about how L2 collocation knowledge interrelates with other types of lexical knowledge and overall proficiency in the second language. A final point is that L2 collocation knowledge has largely analysed linguistic products rather than learning processes, so that no clear picture has emerged of how learners themselves address the problem of developing their collocation knowledge in a second language. These four points—the focus on collocation types, collocation production and recognition, the interrelationship of collocation knowledge to lexical knowledge and overall proficiency, and the learning of collocations—form the basis of the discussion in this section.

2.5.1 The focus on collocation types

Previous investigations have tended to limit their focus to particular collocation types. The extreme examples of this focus on collocation type are the studies by Zhang and Gitsaki. These two studies retrieved a very large number of collocation types from learner corpora and measured L2 collocation knowledge by the accurate production of such types. While other studies focused on a single particular type of collocation such

as V + N combinations, previous research has by and large seen L2 collocation knowledge as located in the accurate control of specific adjacent 2-word linguistic forms. This focus has been accompanied by a concern with examining mis-collocations in order determine how collocationally proficient learners are. In turn, this has resulted in a rather one-sided, L2 deficit-driven view of the development of L2 collocation knowledge where erroneous linguistic forms have by default functioned as the operational construct. If we want to move beyond a preoccupation with collocation error, we need to question how we might elicit and analyse learner productions of language in a different way. It might, for example, be possible to identify a set of target items in written productions and look at the collocation environments of such items across different individuals. That kind of focus on collocation spans would offer us a different entry point for beginning to understand L2 collocation knowledge. It might also be possible to elicit learner productions of collocations using decontextualized measures.

2.5.2 Collocation production and recognition

Previous research has been predominantly concerned with L2 collocation production, not recognition. We noted in our discussion of Biskup's study that collocation recognition is assumed to be unproblematic for L2 learners, but there is very little evidence in Biskup's investigation or any other to show that such a claim is warranted. Although the results from both Howarth's and Nesselhauf's studies indicate that verb knowledge is critical in determining successful V + N collocation use, no research has been conducted into how well the individual verbs and nouns of particular V + N collocations are known. Sentence-level and essay-based investigations have moreover suffered from the very small number of items that they have worked from or been able

to elicit from different individual learners. To take just one example, Bahns and Eldaw's study is widely cited as one of the major works into L2 collocation knowledge. Yet, we need to underline the fact that B&E used just 15 V + N combinations. In fact, it is really only the experimental work of Read and Mochizuki that has attempted to deal with the question of collocation recognition. These experimental studies tried to develop particular decontextualized measures of a rather large number of items for estimating collocation knowledge. Read and Mochizuki demonstrate that it is possible to obtain databases of L2 collocation knowledge without restricting the focus to sentence- or text-level elicitation. Here, the assumption has been that, if learners can match the separate elements of a collocation, they know the collocation, and that such matching is an adequate measure of L2 collocation knowledge. So, despite moving beyond a preoccupation with error, the experimental measures from previous research could be developed into more sensitive measures of collocation recognition that test larger numbers of items in a short time. Just how to do this remains as yet unclear, but it is obviously an area where further experimental work could be done.

2.5.3 Collocation knowledge and other types of lexical knowledge and proficiency

Previous studies have taken different positions as to how L2 collocation knowledge relates to other types of lexical knowledge and overall proficiency in the second language. Some have concluded that collocation knowledge is a decidedly unstable form of lexical knowledge, while others have found that the development of L2 collocation knowledge is limited to advanced learners. In terms of English proficiency, most studies have looked at advanced learners of English, while just five of the previous studies have sought to measure collocation knowledge at intermediate and lower levels. It would therefore be useful to test L2 collocation knowledge where some independent

measure of L2 proficiency can be compared. It would be equally useful to test learners of differing proficiency if we want to get a clearer idea of whether the development of L2 collocation knowledge is restricted to advanced learners or not.

2.5.4 Learning collocations

Previous research has generally overlooked considering in any great detail how learners themselves address the challenge of developing their L2 collocation knowledge. Read used interviews to validate the word associates test and arrived at some interesting insights as to how learners make the kind of collocation links tested by the WAT. Yet, these insights were limited to the experimental context and did not attempt to understand how learners more generally approach the learning of collocations and what vocabulary learning routines and strategies they favour, and why. Other previous experimental work has examined linguistic evidence to advance detailed post hoc reconstructions of how the L2 mental lexicon works. Howarth states, for example, that mis-collocations are sometimes produced because the two nouns are "considered to be semantically too dissimilar to be easily confused in the lexicon, and there exists a verb which collocates with two of the nouns in the hypothetical set" (Howarth, 1998, p.182). This suggests a very orderly type of L2 lexicon in which lexical items are organised in semantic sets, but there is simply not enough direct evidence at all in the previous research to warrant such claims. So, a final area of concern for future research is to try to model the L2 mental lexicon in ways that can more clearly account for the development of collocation knowledge.

2.6 Conclusion

In this chapter, we have taken a detailed look at previous research into L2 collocation

knowledge. This literature review has allowed us to identify different limitations and areas of concern. Discussing these issues enabled us, in particular, to question whether a preoccupation with collocation error is the best way to conduct such research. We have decided that it is largely inappropriate. We have also raised questions as to the limited number of examples that previous research has generalized from. We have the sense that contextualized measures could be combined with decontextualized experimental measures of collocation recognition and production, if the research is to be more informative. We are as yet uncertain as to how that combination of instruments might be best managed, just as we do not yet know what kind of data might be best collected and analysed for our research purposes. We will therefore pursue these questions in more detail in the experimental work presented in the following eight chapters.

Chapter 3

Exploring the Collocation Environments of Lexical Verbs

3.0 Overview

In this chapter, I start by considering the complex nature of previous L2 collocation research—complex in the sense that many studies have, so to speak, mixed and matched different interpretations of collocation in investigating L2 collocation knowledge. I develop this notion of complexity with specific reference to one large-scale inventory analysis study (Gitsaki, 1996). I consider how a different, simpler approach might be taken to analyzing a small corpus of essays. This alternative approach restricts itself to the textual interpretation and focuses on exploring the collocation environments of lexical verbs. The design of this first experiment is explained, and the results of the analysis reported, before I discuss how learners tend to collocate high frequency and academic lexical verbs in differing ways. I conclude this chapter with a critical evaluation of the following fundamental issue: Is a textual approach necessarily the most appropriate for investigating L2 collocation knowledge?

3.1 Introduction

In Chapter 1 we considered three major interpretations of collocation in L1 research—the phraseological, the typological and the textual—and found that the corpus-based textual approach has somewhat resisted reducing collocation to adjacent two-word lexical combinations. Sinclair's work in particular suggested that lexical structures larger than the unit of the word are needed for tracing the distribution of meaning across text. His analysis of **budge** (Sinclair, 2004b) showed that the collocation environments of particular lexical items extend further than their immediate neighbouring left and right neighbours in a text.

Different studies of second language collocation knowledge were reviewed in Chapter 2, several of which also used corpora for exploring learner production of collocations. These L2 textual investigations could be divided into two sub-categories according to what specific focus was taken. Some studies exploited essay corpora corpus for the narrow analysis of particular collocation types such as lexical V + N collocations (i.e., Howarth, 1998; Nesselhauf, 2003) and delexical V + N combinations (i.e., Chi, Wong & Wong 1994). Other investigations conducted large-scale inventory-based analyses of prompted essay texts (e.g., Zhang, 1993; Gitsaki, 1996), where the goal was to parse L2 written productions for different kinds of grammatical and lexical collocation. Both sets of investigation worked from pre-established types of collocations in order to measure the accuracy of L2 collocation production in text.

Gitsaki's study with Greek learners of English demonstrates well the complex character of many previous L2 collocation studies (Gitsaki, 1996). She put forward two hypotheses for investigation:

1. There are stable patterns of development of collocational knowledge across language proficiency levels.
2. There are stable patterns of development of collocational knowledge within language proficiency levels.

The following division between lexical and grammatical collocation was used to inform the analysis of the essay data:

> ...if a collocation is lexicalised, i.e., if the combination of an open class word (verb, noun, adjective, adverb) and a preposition or another open class word is used as single word, e.g., **to do one's homework, to depend on, strong in**, then it is a lexical collocation. If the collocation is a combination of an open class word (verb, adjective, noun, adverb) and a clause, infinitive, gerund, or preposition, then it is a grammatical collocation, e.g., **enjoy + V-ing, want to + infinitive**. (Gitsaki, 1996, p.52)

This created an inventory of 37 types of collocation, of which 11 were lexical, and 26 grammatical (see Appendix 2.1 for an example of each type). The study used essay

writing, translation, and blank-filling to investigate the development of L2 collocation knowledge within three levels of proficiency (N=275): low, mid, and high. Gitsaki further measured language proficiency with six dependent variables (Holistic Rating, Target-Like Use of Articles, Lexical Density, Words per T-Unit, Error-Free T-Units, and S-nodes per T-Unit). To identify the consistency of L2 collocation knowledge development within and between groups, accurate use of the collocation types in each essay was analysed, and, with the help of various statistical measures, Gitsaki concluded that both her hypotheses held.

Although the Gitsaki study is perhaps an extreme example of the mixing and matching of interpretations, types and measures, we may observe that, compared to L1 research, there is something of a complexity in the way that previous L2 corpus-based research has interpreted collocation—and this complexity leads to problems in making clear sense of the results. Two main problems can be, for example, identified in Gitsaki's study. The first is that a wide range of *a priori* collocation types was used in the data analysis. This means that the analysis focused on a broad range of structural phrases rather than the varied co-occurrence of different lexical items. The effect of such an analysis is to mix the colligational with the collocational and to lose sight of the purely lexical characteristics of L2 collocation production. A second problem is that the text lengths of individual essays were not standardized in Gitsaki's study, so it was uncertain what the occurrence of a particular collocation type in a particular written production might in fact point to. Clearly, there is a need to keep things simpler if we are to observe the development of L2 collocation knowledge more clearly.

One way to keep things simpler and clearer is to avoid mixing the textual with the typological. If we focus, for the time being, just on developing a textual interpretation of L2 collocation knowledge, we may observe the wider collocation environments of

particular lexical items and their patterns of lexical co-occurrence. As this could prove a useful way, first, to understand L2 collocation knowledge and, second, to inform the design of later experimental work, we set out, in this chapter, on such a narrower, and more open-ended, inquiry.

3.2 Method

For this experiment, I used the essays of 38 first-year undergraduates as the written corpus. The students belonged to the same English writing class (Humanities A) and had been placed in this intermediate class on the basis of the University of Tsukuba English placement test (which has been shown to give a reliable initial measure of general English proficiency). The students individually chose a topic they wished to research and then spent time discussing, drafting, and revising their essays over a number of weeks in the first term of the academic year (one 75-minute lesson a week). I did not intervene in any way in the writing of the essays, so the essays represent a general picture of the students' uncorrected written proficiency. The students selected a wide range of topics to write about, such as *Brain Death*, *The Peruvian Japanese Hostage Crisis*, *Making Museums User-Friendly*, and *Folklore Studies*, so they did not use any common reading sources as starting input. This free selection, however, left the corpus somewhat varied in the type of lexis that the students produced.

The student essay corpus that I intended to use came to a total of about 20,000 words, and individual essays varied in length from 500 to 1300 words. With such a variety of written productions, I decided to take the following steps so that I could address the two problems noted earlier with Gitsaki's study. First, in order to create standardized texts for comparison, only the first 200 words from each of the 38 essays were chosen for analysis. The second decision was to divide the essays into groups of proficiency according to a sole

measure of accuracy rather than to apply a full battery of six measures of accuracy as Gitsaki had. A third decision was to refrain from looking for *a priori* types of collocation. Rather, it seemed more useful to tag the student corpus for lexical verbs and to look at which verbs were common to all essays. This would provide insights into specific patterns of lexical use and co-occurrence. The results from analysing the tagged verbs could also be used to explore collocation use by examining the collocation environments of (a) verbs common to all essays, (b) particular high frequency verbs, and (c) and particular low frequency, or academic, verbs. Collocation environments here refers to lexical items collocating with a particular lexical verb within a span of four words left and right of the node (Sinclair, 1991, p.175).

The 38 two-hundred word samples were separated into two groups on the basis of T-unit analysis. T-unit stands for "minimal terminable unit" (Hunt, 1966, p.737) and was originally used by Hunt to refer to "one main clause plus whatever subordinate clauses are attached to that main clause" (Hunt, 1966, p.737). Thus, according to Crookes, "*Mary hit John* is one T-unit, and *Mary hit John, but she is my best friend* is two T-units" (Crookes, 1990, p.184). For the present study, T-unit is taken to mean any clause that can stand alone in a grammatically correct fashion.

T-unit analysis involved marking all clauses in each 200-word sample as error-free or error-inclusive. Two groups were created on the basis of the mean of the number of error-free T-units in each essay (mean = 15.58, SD = 0.95), as shown in Tables 3.1 and 3.2 below. (It should be noted that the there are problems of near overlap between the two groups in making such a cut-off point with the low standard deviation.) Student essays which scored equal or lower than 15.58 were assigned to the low proficiency group (n=20); and student essays that scored higher than 15.58 were assigned to the high proficiency group (n=18).

Table 3.1 T-unit results for the higher proficiency group

Error-free T-units	Words in error-free T-units	Error-inclusive T-units	Words in error-inclusive T-units
31	190	3	10
26	149	9	51
25	181	2	19
25	157	8	42
24	167	3	33
22	120	12	80
22	144	5	56
20	132	9	68
20	150	7	50
19	114	8	86
19	167	3	33
18	130	10	70
17	114	10	86
17	110	6	90
16	145	7	55
16	95	15	105
16	93	11	107
16	93	14	107

Table 3.2 T-unit results for the lower proficiency group

Error-free T-units	Words in error-free T-units	Error-inclusive T-units	Words in error-inclusive T-units
15	108	9	92
15	91	14	109
15	85	15	115
14	84	16	116
13	82	14	118
13	99	13	101
13	79	12	121
12	77	16	123
12	113	11	87
11	98	11	102
11	51	15	149
11	86	12	114
11	69	17	131
9	68	16	132
9	63	17	137
9	55	15	145
8	56	10	144
8	56	13	144
8	41	17	159
6	36	18	164

Each text was marked up with the tags <verb>, <aux verb> and <mod verb> for analysis with *Wordsmith 2.0* (Scott, 1998). Auxiliary verbs and modal verbs were excluded from the subsequent analysis of lexical verbs in the data. This decision was taken for the following reasons. Firstly, auxiliary verbs would offer little substantive lexical information with such a small corpus; secondly, analysis of the limited number and type of tagged modal verbs (*can, can't, could, may, should, shouldn't, must*) revealed no significant difference between the two assigned proficiency levels. Lists of concordances of verbs were then generated for each essay, and the total number of different lexical verbs for each group was calculated.

3.3 Results

The results for the comparison of groups are reported first. This is followed by the results for (a) lexical verbs common to all essays, (b) a small set of high frequency verbs, and (c) a small set of particular low frequency, or academic, verbs are given.

3.3.1 Comparison of groups

The lower proficiency group generated a total of 150 different lexical verbs (mean = 14.85, sd = 4.15), and the higher proficiency group produced 131 (mean = 15.89, sd = 3.01). A t-test showed no significant difference between the two groups.

3.3.2 Lexical verbs common to all essays

Taken together, the combined sample produced a total of 227 different lexical verbs. In order to see patterns of frequency, the lexical verbs from the whole sample were marked for presence (1) or absence (0) in all of the 38 essays. The complete results of this analysis are shown in Appendix 3.1. Only four lexical verbs were found to be

common to all samples, namely **be**, **have**, **want**, and **think**. I then chose **want*** for further analysis. The concordance for **want*** produced 41 concordance lines, as shown in Table 3.3 below.

Table 3.3 Concordances for want*

Jyoho-Gakurui, for example, they	**want**	Their major is concerned
Japanese drivers and I	**want**	them at least to complete the
as a same Japanese,	**want**	them to take a good rest
their life as they	**want**	They want to get
going to be, I	**want**	to be. Some of
professional baseball player "You	**want**	to become an ordinary
reasonable, though I don't	**want**	to believe that there
salary paying job. They	**want**	to get many kinds
of modern life. They	**want**	to get more interesting
us. The characters don't	**want**	to have hard relationships
America and Mexico. I	**want**	to know about there
In particular what I	**want**	to know is the
to think that I	**want**	to know more and
freely? I came to	**want**	to know.
life. Of course, I	**want**	to live ideally, but
very interested. So I	**want**	to research about Okinawa.
I	**want**	to research about the
be no hope. So I	**want**	to research about these
I	**want**	to research is about
I	**want**	to research Josaku in
since then. I also	**want**	to research local history.
What I	**want**	to research now is
I	**want**	to research the history of China.
clear reason why I	**want**	to research the history
the topics that I	**want**	to research. I'm interested
bad thing. What I	**want**	to say to many
makes funny situation. I	**want**	to see many other
always win the race. I	**want**	to see the minor
although, most of them	**want**	to study and have
study at college. We	**want**	to study at college
to think that I	**want**	to study folklore. Before
I	**want**	to study Japanese folklore
history. Nevertheless, now I	**want**	to study Japanese folklore.
research is Ecology. I	**want**	to think and learn
The topic I	**want**	to think is museum
I	**want**	to think more about
for a while. I	**wanted**	to do something for
had curiosity that I	**wanted**	to know a lot
historians that they had	**wanted**	to know everything about
was natural that I	**wanted**	to study history. Nevertheless
folklore. Before that I	**wanted**	to study Japanese history

In Table 3.1, the concordances include four words to the left, and four to the right. (Where the concordance starts 'in', the concordance comes from the opening sentence of a particular essay.)

Finer analysis of the 41 concordance lines for **want** was carried out by restricting the collocate distance to one word to the left and right, and two words to the right respectively (Biber, Conrad & Reppen, 1998, pp.51-54), as shown in Tables 3.4 and 3.5 below.

Table 3.4 Most common collocates of want occurring one word to the left and to the right

Left collocate	%	Right collocate	%
I	66%	**to**	90%
They	12%	**They**	5%
don't	5%	**them**	5%

Table 3.4 shows that **want** shows a high co-occurrence with common everyday general English. The immediate left and right collocates are all included in the first 150 words of the Birmingham Corpus (Sinclair & Renouf, 1988). The immediate collocation environment of **want** was found to be colligational (i.e., syntactic), in other words.

The collocation span of **want to** was somewhat different. In Table 3.5, we can see that **want to** collocates mainly in the first person with verbs of mental process and activity, such as **research**, **know**, **study**, and **think**.

Table 3.5 Most common collocates of want to

Collocates	%
research	24%
study	17%
know	15%
think	3%

In Table 3.5, 10 instances co-occur with **research**, eight with **study**, five with **know**, and three with **think**. In turn, these four collocates belong to the 12 most frequent verbs in the corpus (see Appendix 3.1). Moreover, **research**, the most frequent collocate with **want to**, is listed by Nation in the UWL (Nation, 1990).

3.3.3 A small set of high frequency lexical verbs

The collocation spans of a set of four high frequency verbs (**appear**, **believe**, **explain**, and **raise**) were analysed and compared to the sense definitions given in a learner's dictionary (Sinclair, 1990). Results showed that the four lexical verbs were used in the present corpus in common everyday senses with a relatively strong relationship to spoken use, as shown in Table 3.6 below.

Table 3.6 Concordances for four high frequency lexical verbs

caused a flood, a star	**appeared**	in the sky
in detail. Because after human	**appeared**	, it has been very long
death called "brain death" have	**appeared**	Should we accept the new
I think the Japanese doesn't	**believe**	in their religion heartily. When
What kind of religion we	**believe**	in? Christianity, Buddhism, Shinto or
though I don't want to	**believe**	that there are many students
thinking. It seems that they	**believe**	in their religion heartily. How
through studying archaeology. Let me	**explain**	why I have become interested
First of all, let me	**explain**	the places which impressed me.
In the beginning, let me	**explain**	about brain death briefly for
members were killed although they	**raised**	the white flags up. Perhaps,
would not kill them who	**raised**	the white flags. However, the

3.3.4 A small set of low frequency lexical verbs

The University Word List (Nation, 1990) was used to identify a small set of low frequency academic verbs in the student corpus (see Table 3.7 below).

Table 3.7 Occurrences of UWL verbs

Occurrences	Academic verb
12	research
3	criticize
2	impress, maintain
1	adapt, clarify, conclude, consist, decline, devote, dominate, emerge, emphasize, establish, evolve, import, inherit, insist, lecture, maintain, occur, participate, pollute, pursue, reject, reveal, select

The collocation spans of these academic verbs were then examined. Table 3.8 below shows the results for the lexical verb **research**.

Table 3.8 Immediate environment collocates for RESEARCH (closest noun phrase to the left and right)

Nearest noun phrase to the left	Node	Nearest noun phrase to the right
I	**research**	agriculture
I	**research**	American military bases on Okinawa
I	**research**	I
I	**research**	I
I	**research**	josaku (=*fortress administrations*)
I	**research**	Latin America and Mexico
I	**research**	local history
I	**research**	Okinawa
I	**research**	Okinawa
I	**research**	the history of China
I	**research**	the history of China
I	**research**	the media culture
I	**research**	the whole problem
I	**research**	these problems
I	**research**	these questions
I	**research**	what kind of problems

The 16 right collocates were classified into the following patterns:

- **research** + *specific place* (4 occurrences)
- **research** + *history/culture* (4 occurrences)
- **research** + *problem/question* (4 occurrences)
- **research** + *self* (2 occurrences).

Two of these collocate categories (**research** + *history/culture* and **research** + *problem/question*) showed evidence of a move away from concrete here-and-now collocation environment towards an academic use.

The collocation spans of the other academic lexical verbs (i.e., the nearest noun phrase to the left and right of the particular verb) are presented in Table 3.9 below, where UWL noun collocates are highlighted in bold.

Table 3.9 Immediate environment noun phrase collocates for UWL verbs (with UWL items highlighted in bold)

Nearest noun phrase to the left	Node	Nearest noun phrase to the right
no knowledge and experience	**criticize**	this event
biographies	**criticize**	the **traditional** way of thinking
this lightness	**criticize**	heavy literature
natural beauty of the Lake District	**impress**	I
the places	**impress**	one of these places
I	**impress**	May 15th of this year
it	**maintain**	the shape of pyramids
two schools	**maintain**	each demand
the Japanese **economy**	**adapt**	this situation
learning	**clarify**	an **ethnic** history
one thing	**conclude**	a national opinion
authority of king	**decline**	Egyptian people
Christians and Muslims	**devote**	their religion
time and information	**dominate**	us
these days	**emerge**	many demands for **constitution**
some of Japanese movies	**emphasize**	the **mental aspect**
our **culture**	**evolve**	magazines
Japan	**import**	much food
it	**inherit**	public life
many products	**insist**	their quality and low-use **energy**
our teacher	**lecture**	those
these **phenomena**	**occur**	it
one of the school **regulations**	**participate**	club
many rivers or seas	**pollute**	the **area** of deserts
the book	**pursue**	the changes of astronauts' insides
our life	**reject**	our nonsensially rich life-**style**
purpose	**reveal**	magic tricks
I	**select**	one of them

Table 3.9 shows that 12 of the 30 UWL verbs featured collocation environments that included an UWL item in the noun phrase collocates.

3.4 Discussion

Although the essay database could be divided into proficiency groups by T-unit analysis (essentially a grammatical measure of proficiency), there was no basis for claiming a difference in lexical proficiency between the two groups of students. There is some reason to believe that lexical accuracy itself may not necessarily be a valid indicator of level since higher-level learners may make more lexical errors through greater active experimentation and use of vocabulary than lower-level learners. Although the issues of lexical accuracy and lexical proficiency need greater consideration, such an elaboration is outside the scope of this replication. If claims about the development of L2 collocation knowledge are to be limited to the broad band of intermediate proficiency that these 38 students have, they must be treated as general observations rather than significant findings. I will therefore limit the discussion to three points that focus on observable collocation differences between general everyday lexical verb collocations and academic lexical verb collocations. These points concern the collocation environments of the small set of high frequency lexical verbs, the collocation environments of the small set of low frequency academic lexical verbs, and the structural classification of academic verb noun phrase collocates.

In the presentation of results, we noted that the collocation environment of **want** showed both a typical high-frequency environment and an emerging specific low-frequency collocation restriction. If we see the high frequency of **want** as an interlinguistic indication of writer intention and concern with mental process, it becomes clear that, at this level of proficiency, with this particular essay writing task,

want and its collocates seem to be acting as an explicit organiser of the written discourse. At the same time, such a high frequency use of this lexical verb would seem unusual in more advanced academic writing.

The analysis of the collocation environments of the four highly frequent lexical verbs further showed that the collocates of **appear, believe, explain,** and **raise** tended to be strongly restricted to the here-and-now world of concrete reality. The verb **appear** was used in the sense of ***materialize*** or ***come into use***, but not as an authorial distancing device for weakening claims in an argument (Swales & Feak, 1994, pp.86-87). The second verb, **believe**, was employed in the sense ***of being certain that something exists***, but not in the sense of ***reporting positions that others maintain***, and **explain** was used in the sentence header ***Let me explain*** to explain the writer's own point of view rather than to ***report or quote the ideas of other people*** (two out of three occurrences). Finally, the lexical verb **raise** occurred in the physical sense of **surrender** (by raising a white flag), but not in the more abstract sense of ***bring a question or objection up in an argument***. These examples appear to illustrate how L2 collocation use seems to be strongly oriented at the intermediate level to the most frequent here-and-now concrete senses of particular lexical verbs, which, at a higher level of proficiency, will tend to have other important functions in the organisation of academic writing.

Where academic verbs are used by intermediate learners, several interesting features can be tentatively observed. First, it seems that noun phrase collocates show greater structural complexity when they co-occur with UWL-type verbs; similarly, single pronominal collocates, which were seen to be highly frequent with **want** (cf. Table 3.4), are much more limited in their co-occurrence with UWL-type verbs. These two points suggest a shift away from the use of here-and-now concrete everyday senses

that we noted with the highly frequent lexical verbs towards more abstract material things, places, relations, and processes in the collocation environments of academic lexical verbs. This shift seems to change according to the frequency of the verb lemmas in the corpus; that is, the more frequent the academic verb is in the corpus, the greater the range of abstract collocates in its immediate environment.

By examining the concordances for these verbs and identifying the collocates in terms of the nearest noun phrase to the right and to the left, we can begin to get an initial trace of how the shift from informal everyday use towards more formal academic use is patterned. From a syntagmatic point of view, the noun collocates of academic verbs can be structurally classified into two groups of strong adjectival pre-modification and strong nominal post-modification (see Table 3.10 below). In the multi-word noun phrases, one element of the noun phrase comes from common high frequency everyday vocabulary and is combined with a less frequent word (shown in bold) to realise a more abstract concept, e.g., the Japanese **economy,** our **culture,** one of the school **regulations,** these **phenomena**.

Table 3.10 indicates that, in the strong adjectival pre-modification group, the more common element also tends to be adjectival in such cases. However, in the strong nominal post-modification group, the adjectival development of the noun phrase collocates seems to be restricted, for instance: **authority of king, the shape of pyramids, the area of deserts**. We can also note: **the traditional way of thinking, the natural beauty of the Lake District**. These last two examples indicate that a high frequency principle seems to be applied to the more frequent noun, which in turn is itself pre-modified by a less frequent adjectival: **the traditional** (=less frequent adjectival) **way** (=more frequent nominal) **of thinking** (less frequent

nominal), **the natural** (=less frequent adjectival) **beauty** (=more frequent nominal) **of the Lake District** (=less frequent nominal).

Table 3.10 Structural classification of UWL verb noun phrase collocates

Type of noun phrase	Realisation
Simple noun phrase (single word)	I, it, learning, biographies, Japan, us, magazines, purpose, those, club
Group 1: Strong adjectival premodification	
Deictic + Noun	the places, this lightness, this event, this situation, these days, these phenomena, the book
(Deictic) + Classifier + Noun	heavy literature, Egyptian people, their religion, much food, public life, magic tricks, our life, our teacher, our culture, the Japanese economy, an ethnic history, a national opinion, the mental aspect
Numerative + (Deictic) + (Classifier) + Noun	one thing, each demand, two schools, many products, one of these places, one of them, some of Japanese movies, one of the school regulations
Numerative + Noun + Preposition + Noun	many demands for constitution
Classifier + Adverbial + Epithet + Noun	our nonsensially rich life-style
Group 2: Strong nominal post-modification	
(Deictic/Numerative) + (Classifier)+ Noun + Preposition (+Deictic) + Noun	authority of king, the shape of pyramids, the natural beauty of the Lake District, the traditional way of thinking, May 15th of this year, the changes of astronauts' insides, the area of deserts
Group 3: Coordinated nominal groups	
(Classifier) + Noun + AND/OR + (Classifier) + Noun	no knowledge and experience, time and information, many rivers or seas, their quality and low-use energy, Christians and Muslims

The above observations thus tend to suggest a tension of frequency between the different elements in the respective collocate group, as Sinclair notes: "Most normal text is made up of the occurrence of frequent words, and the frequent senses of less frequent words" (Sinclair, 1991, p.113). Conceptualising this tension in terms of frequency seems, then, an appropriate way of understanding the combination of the different elements.

We may also hypothesize an initial 3-stage colligation frame for the development of noun phrase collocates with less frequent, academic lexical verbs:

Stage 1: Strong adjectival pre-modification (base pattern: Adjectival + Noun)
Stage 2: Strong nominal post-modification (base pattern: $Noun^1$ + $Noun^2$)
Stage 3: Combinatory elaboration of Stages 1 and 2 (base pattern: Adjectival + $Noun^1$ of $Noun^2$.

In fact, Gitsaki lists the two patterns of Adjectival + Noun and $Noun^1$ + $Noun^2$ among the 11 lexical collocation types that she used (see Appendix 2.1), but the list does not account for the combination of the two types into the Stage 3 pattern that we have noted here (Adjectival + $Noun^1$ of Noun). We have some evidence from the present experiment that the development of noun phrase collocate environments is somewhat lexically more fine-grained than a taxonomy of lexical collocation types points to.

To sum up, the goal of this modified replication was to work through a set of procedures simplified from Gitsaki's research in order to explore L2 collocation development. Gitsaki was able to confirm both her hypotheses, namely:

1. There are stable patterns of development of collocational knowledge across language proficiency levels.
2. There are stable patterns of development of collocational knowledge within language proficiency levels.

In the modified replication, I was unable to address the first hypothesis because I had data from just one class group where the T-unit analysis indicated no significant difference in proficiency. This study was therefore restricted to one broad band of

intermediate proficiency and was based on an analysis of a standardized sample of the first 200 words from 38 student essays. The experiment further limited its focus to different lexical verbs and their collocation environments.

To make sense of the data, frequency was used as an initial sorting device. This restricted the choice of lexical verbs for further analysis. The greater the number of concordances, as with **want** and **research**, the richer (and more confident) the collocation insights were. The more limited the number of concordances for each verb, as with the academic verb selection, the more I tried in the analysis to systematise the surrounding collocation environment. The collocates themselves here offered richer lexical information than the limited number of examples with each individual verb.

Although I chose not to look for *a priori* collocation types in the analysis of the data, it was nevertheless tempting at times to do this in order to make systematic sense of the data. This made me realize the attraction of imposing pre-determined patterns on a set of data, but such an approach would have misdirected the lexical analysis and missed data-driven patterns. Simply put, it would have led to a complex interpretation and unclear results. Nevertheless, we must underline a persistent weakness with a corpus approach to investigating L2 collocation knowledge. It is simply very difficult to retrieve enough examples of particular lexical items and their collocation environments for generalizable claims to be made. This was a recurrent limitation noted in Chapter 2 in the review of previous L2 studies. From a 1-million word corpus, the Hong Kong study (Chi, Wong & Wong, 1994) identified, for example, just 167 delexical V + N combinations for analysis (12 combinations with **have**, 16 with **do**, 44 with **make**, 46 with **take**, and 49 with **get**). This suggests that following a solely textual interpretation of collocation may run into logistical difficulties.

The question now is how to continue. The tension between frequent and infrequent vocabulary in collocations seems central. However, it also seems important to investigate further learners' knowledge of academic lexical verbs, before we look at their collocation knowledge of such verbs. It may be quite possible to explore knowledge of academic lexical verbs without using a corpus if we can devise a suitable decontextualized measure of verb knowledge. There are three main reasons for believing this is a sensible path to follow. First, we have evidence from this experiment that learners' productive use of academic verbs is quite limited. Second, we do not know whether learners' recognition of such verbs will be similarly limited or not. Finally, we need to establish a greater bank of baseline data if we want to have solid foundations for designing later L2 collocation knowledge experiments.

3.5 Conclusion

This chapter has presented a modified replication of a large-scale inventory analysis approach to corpus data. I started out by questioning whether I should focus on collocation types or collocation environments. I chose to do the latter so that I could recover instances of lexical co-occurrence and avoid taking a complex approach to understanding L2 collocation knowledge. I first focused on the collocation environments of high frequency lexical verbs and then examined the interaction between academic lexical verbs and different noun phrase collocates. The analysis could not be based on statistically significant findings because the number of realizations that I found was very limited. However, the observations that I did make suggested a tension between high frequency and low frequency lexis in the L2 collocations that were used. My observations also suggested that learners' productive knowledge of academic lexical verbs is limited. This made

me consider what focus I should take in the next experiment. I have decided to opt for designing and using a decontextualized measure of academic verb knowledge, so that I can later develop appropriate instruments for investigating L2 collocation knowledge. In the next chapter, then, we step back from a textual interpretation of collocation and explore how we can measure learners' knowledge of academic lexical verbs in an appropriate, decontextualized way.

Chapter 4

A Preliminary Exploration of Lexical Verb Knowledge

4.0 Overview

In the previous chapter, we used a small corpus of student essays to start to explore the development of L2 collocation knowledge. We limited the analysis to the collocation environments of high frequency and academic lexical verbs. We found that the high frequency verbs were mostly used in common everyday collocations and that the academic verbs tended to co-occur with UWL noun phrases. The analysis had a tightly controlled focus because we wanted to develop a clear textual understanding of second language collocation production. That let us see that the combination of frequent and infrequent lexical items in L2 collocation production is of potentially great importance. Yet, while we have good grounds for wanting to explore this area further, we must leave such exploration aside for the time being. The reason is that we need to gather more compelling evidence of learners' knowledge of academic lexical verbs first. In this chapter, we will consider how to do that, so that we can later investigate the development of L2 collocation knowledge from firmer foundations.

4.1 Introduction

A corpus linguist might wonder what could be more compelling than the authentic evidence gleaned from examining real-world text. This is very much the position that Hoey takes when he states that collocation is more than anything "a textual phenomenon" (Hoey, 1991, p.219). However, we face a major problem in investigating L2 collocation knowledge through text alone. The problem is that we simply do not have at our disposal the multi-million word corpora that are the staple diet of L1 corpus

linguistics. By accepting that an L2 corpus-based study may not provide us with sufficient information, we need to develop a measure that is independent of essay data. The measure of academic lexical verb knowledge that we will develop and evaluate in this chapter will be discrete, selective, and context-independent (Read, 2000, pp.7-13). As we shall see, this measure is discrete in that it aims to measure a particular aspect of vocabulary knowledge. It is also selective because it focuses on particular academic lexical verbs. Finally, it is context-independent because it does not require the learner to refer to context in order to complete the test. It is important to highlight these broad dimensions of the measure at the outset, so that we can keep in mind a clear sense of the different approach that we are going to take in this study. This more experimental approach will allow us to measure learner knowledge of a large set of pre-determined items and also provide us, we hope, with a useful basis for exploring L2 collocation knowledge later. This chapter focuses, then, on the development and evaluation of such a decontextualized measure of academic lexical verb knowledge.

4.2 The Vocabulary Levels Test

There are few standardized tests of vocabulary knowledge available. The two most widely used tests for estimating vocabulary size are Nation's Vocabulary Levels Test (Nation, 1990) and the Checklist Tests ('yes/no' tests) developed by Meara and colleagues (Meara & Buxton, 1987; Meara, 1992; Meara & Milton, 2003). Generally speaking, the VLT is more widely used by teachers than the checklist tests, and it was this one that I was originally more familiar with. In the original format of the test, 18 words are tested at the 2,000, 3,000, 5,000, and 10,000-word levels, as well as in the University Word List range. The UWL is based on coverage rather than frequency of approximately 800 words common across academic disciplines and has been recently

revised and updated as the *Academic Word List* (Coxhead, 2000). Each section of the Nation test contains six sets of 6 different words, grouping nouns, verbs, or adjectives in each set of definitions. Three words per set must be matched with three short definitions. An example set from the UWL section is:

1. configuration
2. discourse _____ **shape**
3. hypothesis _____ **speech**
4. intersection _____ **theory**
5. partisan
6. propensity

Nation does not claim that the test measures total knowledge of any particular item; it simply estimates "learners' receptive vocabulary" (Nation, 1990, p.93) by estimating their ability to match well-known and partially known words with their correct "definitions" (Nation, 1990, p.262). Schmitt defines the VLT more specifically as measuring "threshold meaning knowledge of the target words" (Schmitt, 2000, p.174). The test is easy to administer, relatively time-efficient, and easy to check, which makes the test attractive as a diagnostic tool for both pedagogic and research purposes. It is claimed that the results from the VLT give reliable information about "where learners should be given help with vocabulary learning" (Nation, 1990, p.261). We should further note that very few items are used to establish such estimates: 18 words per section if we focus on the number of responses required. We will return to this issue and other problematic aspects of the test when we look at the results of the present study later in this chapter. We will also draw on insights gained from think-aloud interviews with a small group of students so that we can get a better understanding of what kind of lexical knowledge learners exploit in responding to the VLT.

With these initial caveats about the VLT in mind, I decided to create a modified version of the test with verbs taken from University Word List. A 60-item UWL verb

test was created following Nation's original format and broadly parallelling the research design steps of Beglar and Hunt (Beglar & Hunt, 1999). At the time of designing this experiment, their study was the first attempt to validate Nation's 2000 Word Level and University Word Level Tests (Nation, 1990)[1]. The first stage in their validation was norm-referencing the original four forms for each of the two tests (n = 496 for the 2000 Word Level Test; n = 464 for the UWL test). They found large differences in the means and standard deviations between the different forms of each test. They then examined item discrimination values by calculating item-total correlations. Following Ebel's guidelines for distinguishing between very good, good, and poor items (Ebel, 1965), Beglar and Hunt estimated which items were functioning well (see Table 4.1).

Table 4.1 Item-total correlation in Beglar and Hunt's validation

Correlation ranges	.40 and higher (=very good)	.30 to .39 (=good)	below .29 (= poor)
2000 Word Level (72 items)	46 items (63.9%)	14 items (19.4%)	12 items (16.7%)
UWL Test (72 items)	23 items (31.9%)	22 items (30.6%)	27 items (37.5%)

Other language test researchers suggest > 0.4 to be a good benchmark for well-performing items (Alderson, Clapham, & Wall, 1995, p.82) and point out that questionable items should not be rejected out of hand by statistical analysis alone (Alderson, Clapham & Wall, 1995; Hughes, 1989; Oller, 1979). In other words, poor-performing items need to be closely examined before any decision is made to reject them. Beglar and Hunt, in contrast, used item discrimination analysis for selecting the well-performing items as the core of the two revised forms for either test, and, in doing so, moved away from analyzing further possible reasons for low ID values.

The discrimination analysis allowed Beglar and Hunt to winnow the original 72

items down to 54. Their next step was to equate statistically the two forms of each test so that there were no significant differences in the mean scores, equivalent variances, and equivalent covariances. They also used Rasch analysis to calibrate all 54 items in each test before they re-trialled the revised forms (n=119 for the revised 2000 Word Level Test, and n=113 for the revised University Word test). From this, they conducted extensive statistical analysis and further correlated the results with the TOEFL. Here, they found higher correlations between the UWL forms and the Reading Comprehension and Structure and Written Expression subsections than with other subsections of the TOEFL.

To sum up, although the VLT is one of two major tests of vocabulary size used by both teachers and researchers, it was until recently the focus of relatively little research. Beglar and Hunt were the first to attempt to validate and improve the original forms created by Nation, but the validation did not altogether make clear why items perform well or badly or what types of lexical knowledge the VLT measures. As questions remain about the construct validity of the test, it is important to look critically at how the clusters in the test are constructed, how items perform, and, above all, what types of lexical knowledge learners draw on in doing the VLT. By understanding these points, we will be in a better position, first, to evaluate the VLT as a tool and, in later experimental work, to construct a decontextualized measure of L2 collocation knowledge.

4.3 Research goals

The goals of the study are to examine:

1) which UWL verbs students can recognize;

2) which UWL verbs tend to be first known at lower levels of proficiency, and which

verbs are likely to be later known at higher levels of proficiency;

3) what types of lexical knowledge learners exploit in doing the test.

4.4 Research construct and design

Beglar and Hunt's revised University Word Test Forms A and B consist of two 27-item tests. Form A, for example, includes nine items which are verb-based, listing 18 possible choices. Using those items as a starting point, I separated all the verbs from the revised UWL list and created the 60-item verb test, involving 120 possible choices in total. (This approximates to about 15% of the UWL, and approximately 50% of all UWL verbs.) In order to create simple verb definitions for the distractors, I adapted sense definitions from the Collins COBUILD New Student's Dictionary (HarperCollins, 1997). For each set of six choices, I took care to create sets with as distinct senses as possible, as well as to try to make sure that prefixes and suffixes did not overlap.

As I worked on the sense definitions and started constructing clusters, I found that my attention was moving from the adapted sense definitions to each set of six verbs and, finally, to the choice of answer. For these reasons, I decided to alter the layout of the clusters so that they mirrored the stages in the decision-making process that I had noticed. We can see the changed layout in this example cluster:

28. succeed in doing	**a. contradict**	**28. _____**
	b. reformulate	**29. _____**
29. consider closely	**c. accomplish**	**30. _____**
	d. insist	
30. show an idea is wrong	**e. analyze**	
	f. exert	

In the later NS piloting of the test, as well as in the think-aloud interviews with 10 learners, none of the participants commented about the changed layout of the test, so it is safe to conclude that this change was of no great import or effect.

A more important part of designing the test was adapting sense definitions from

the dictionary for the test. Although this appears to be a straightforward step in the research design, creating such definitions involves a simplification of meaning. We can see this by looking again at the example cluster above. Here, the definition for **accomplish** is the closest to the listed dictionary definition,

> *VERB If you **accomplish** something, you succeed in doing it.*

This may be partly due to the fact that **accomplish** has only one sense listed in the COBUILD dictionary, so that the number of semantic variables is limited. The simplification for this item's definition involves the deletion of an explicit grammatical clue of transitivity ('**something**'). In contrast, the simplified definition for **analyze** ('**consider closely**') involves a much greater deletion of semantic information, which results in a highly 'packaged' adverbial element ('**closely**'); here, the dictionary definition of the verb reads as:

> *VERB If you **analyze** something, you consider it or examine it in order to understand it or find out what it consists of ...,*

whereas the definition of the adjectival '**analytical**' seems to add a further important clue:

> *ADJ **Analytical** skills or methods involve the use of logical reasoning ...,*

which, for reasons of space, was not included in the test definition. The COBUILD dictionary definitions for the verb **contradict** and the adjective **contradictory** demonstrate a similar packaging effect, but this time into both the nominal and adjectival elements in the simplified definition used in the test ('**show an idea is wrong**'):

> ***contradict** 1 VERB If you contradict someone, you say or suggest what they have said is wrong ... 2 VERB If one statement or action **contradicts** another, the first one makes the second appear to be wrong ...*
> ***contradictory** ADJ If two or more facts, ideas, or statements are **contradictory**, they state or imply that opposite things are true ...*

It is clear from this that the simplification of the definitions raises questions of both face and construct validity, which cannot at this point be quantified or analyzed further. In

terms of the UWL verb test created, 'knowing' the verb's meaning therefore seems to involve more than just knowing its explicit meaning; it also appears to involve knowing an implied network of immediate but deleted associations—and a good way to explore this further would be to conduct think-aloud interviews with students as they do the test.

Once the draft 60-item test had been completed, four native speakers checked it independently for confusing or misleading items. The test was revised in the light of their comments (see Appendix 4.1 for the final version). I judged that the test would take 20 minutes to complete with first-year undergraduate students.

4.5 Method

The UWL verb test was administered in English to six intact intermediate level classes (Bio-Resources, Humanities, Japanese Language and Literature, Medical Science, Psychology, and Social Policy and Management). Students were willing to participate, applied themselves to the task, and generally showed a positive interest. A total of 188 students took the test; the responses of seven subjects were discarded, because, a month later, they missed taking the English Proficiency Test which was used as an independent measure of the students' general English proficiency (N=181).

Students were instructed to write the letter of their choice next to the number for each item. If they guessed, they were instructed to add an asterisk after the letter. If they did not know and could not guess, they were instructed to leave the item blank. The guessed category was introduced after an initial small-scale piloting of the test revealed that some students reported guessing if they were not completely sure of their choices. The following imaginary example shows this in more detail. Here, the student has indicated that he/she knows the answer is **c** to Q28, has guessed the answer **e** to Q29 and does not know the answer to Q30:

28. succeed in doing	**a. contradict**	**28. c**
	b. reformulate	**29. e***
29. consider closely	**c. accomplish**	**30.**
	d. insist	
30. show an idea is wrong	**e. analyze**	
	f. exert	

What 'guessed' means more exactly was not clear, so guesses were counted as full responses for the statistical analysis.

However, the small-scale piloting of the test and the introduction of the guessing category made me notice that just analyzing the results quantitatively would perhaps be missing out on gaining some important insights into what learners do. I therefore completed think-aloud interviews with a small group of learners as they took the test. These interviews provided some valuable insights into the act of guessing and how guessing involves learners using different types of lexical knowledge. In the results section below, I will also report findings from 10 think-aloud interviews to shed further light on this issue.

4.6 Descriptive statistics

The means and standard deviations can be seen in Table 4.2 below, together with the reliability co-efficients, and ranges. These compare favourably with Beglar and Hunt's findings. In particular, Cronbach's alpha for the reliability co-efficient is high at .88.

Table 4.2 **Descriptive statistics: UWL verb-test compared to Beglar and Hunt revised UWL forms**

	UWL-verb test	B&H revised Form A	B&H revised Form B	B&H revised Forms A & B
n	181	464	464	464
Items	60.00	27.00	27.00	54.00
Mean	32.54	13.03	13.16	26.18
SD	9.57	5.70	5.53	10.79
α	0.88	0.85	0.84	0.92

4.7 Results

In this section, both quantitative and qualitative results are reported. For the quantitative results, I first look at item-total correlations. This offers a useful comparison with Beglar and Hunt's study and gives us a sense of the overall performance of the test (Research Question 1: Which UWL verbs do students recognize?). I then present the results for a high-performing group and a low-performing group so that we can compare two distinct groups of proficiency (Research Question 2: Which UWL verbs tend to be first known at lower levels of proficiency, and which verbs are likely to be later known at higher levels of proficiency?). I conclude the quantitative results by reporting correlations with overall English proficiency are reported. As far as qualitative results are concerned, I summarise insights from the think-aloud interviews so that we can address Research Question 3: What types of lexical knowledge do learners exploit in doing the test?

4.7.1 Item-total correlations

Item-total correlations for the UWL-verb test yield solid results (see Table 4.3 below). According to the guidelines adopted from Ebel (1965, p.267), namely:

.40 and higher	very good items
.30 to .39	reasonably good items possibly subject to improvement
.20 to .29	marginal items in need of improvement
below .19	poor items which need to be revised or eliminated

only two items show themselves to be poor (Items 32 and 38), although a further 14 fall into the marginal category.

Table 4.3 Overview of item-total correlations in the UWL-verb test

.40 and higher (20 items)	.30 to .39 (24 items)	.20 to .29 (14 items)	below .19 (2 items)
4, 11, 13, 14, 17, 19, 21, 24, 26, 28, 31, 33, 36, 39, 40, 47, 48, 50, 53, 54	2, 3, 8, 9, 15, 18, 22, 25, 27, 30, 37, 41, 42, 43, 44, 45, 46, 49, 51, 52, 56, 57, 58, 59	1, 5, 6, 7, 10, 12, 16, 20, 23, 29, 34, 35, 55, 60	32, 38

In the case of Item 32, it appears that the weakness may lie in the reduction of the verb from **function as** to **function,** making the item a highly difficult one.

31. make other people aware of	**a. devote**	**31.**
	b. function	**32.**
32. have the purpose of	**c. overlap**	**33.**
	d. reveal	
33. guess	**e. speculate**	
	f. utilize	

In contrast, with item 38, it seems that the problem may result from the familiarity of **select**, making the item a very easy one:

37. be the cause or basis of	**a. focus**	**37.**
	b. incorporate	**38.**
38. choose	**c. select**	**39.**
	d. acquire	
39. repeat the exact words of	**e. underlie**	
	f. quote	

In both sets, the two other items have very good or good item-total correlations: items 31 (.46) and item 33 (.52); item 37 (.37) and item 39 (.42). The individual item-total correlations are shown in Table 4.4 below, sorted from highest to lowest.

Table 4.4 Individual item-total correlations in the UWL-verb test

#	Item	ID value	#	Item	ID value	#	Item	ID value
33	speculate	.52	30	contradict	.39	8	transform	.32
13	inspect	.51	2	expose	.38	49	devise	.32
21	shrink	.51	18	participate	.38	59	specify	.31
4	evaluate	.49	22	alter	.38	9	assume	.30
31	reveal	.46	41	summarise	.38	6	absorb	.29
17	indicate	.45	52	identify	.38	10	supplement	.29
47.	investigate	.45	37	underlie	.37	55	institute	.29
11	revise	.44	51	converge	.37	23	formulate	.28
48	unify	.44	25	correspond	.36	35	minimise	.28
53	hypothesise	.44	45	facilitate	.36	60	manipulate	.28
14	accumulate	.43	15	emphasise	.35	5	publish	.26
26	clarify	.43	43	vary	.35	29	analyse	.26
19	interpret	.42	46	stimulate	.35	1	rely	.24
24	criticise	.42	57	neutralise	.35	7	restrict	.24
36	derive	.42	3	attain	.34	16	fund	.24
39	quote	.42	27	justify	.34	20	structure	.24
50	assess	.42	42	modify	.34	12	design	.22
28	accomplish	.41	44	require	.34	34	affect	.20
40	conduct	.41	58	emerge	.34	32	function	.15
54	refute	.40	56	enumerate	.33	38	select	.15

4.7.2 Differences in results for a high group and a low group

To explore the second research question as to which UWL verbs may tend to be first known at lower levels of proficiency, and which verbs are likely to be later known at higher levels of proficiency, the next step in the analysis involved looking at the cumulative percentage in frequency distribution and determining the cut-off scores for the top and bottom quartiles. The cut-off points were set at >39 for the top group (74%-100%), > 26 <=39 for the middle group (28%-73%), and <=26 for the low group (0-27%). From this, three groups were created, and a chi-quare test was run for each verb to determine significance levels ($p < 0.05$).

The results are displayed in Table 4.5 for the high and low groups only, so that the correct and incorrect responses for two distinct groups of proficiency can be compared. In Table 4.5, we may note the comparatively high p values for both **select** (.458) and

function (.180), as well as their corresponding non-significant X^2 values (1.56 and 3.43, respectively). The table shows that significant differences between the two groups are found for all verbs except the few most frequent ones.

Table 4.5 Comparison of chi-square results for the low and high groups

Item	Raw percentage (correct)		X^2 *df* 96	*p*	Item	Raw Percentage (correct)		X^2 *df* 96	*p*
	High (n=47)	Low (n=50)				High (n=47)	Low (n=50)		
select	97.9	94.0	1.56	.458	identify	89.4	46.0	20.44	.000
function	25.5	12.0	3.43	.180	emphasise	89.4	50.0	20.59	.000
rely	100.0	86.0	8.79	.012	conduct	66.0	20.0	21.06	.000
analyse	83.0	54.0	9.45	.009	alter	76.6	36.0	21.10	.000
design	72.3	42.0	9.49	.009	enumerate	40.4	4.0	21.33	.000
fund	95.7	74.0	9.52	.009	expose	89.4	44.0	22.14	.000
assume	83.0	54.0	9.86	.007	clarify	72.3	24.0	22.71	.000
affect	97.9	84.0	9.90	.007	derive	76.6	28.0	23.02	.000
structure	57.4	26.0	10.46	.005	criticise	70.2	26.0	23.57	.000
absorb	76.6	44.0	10.88	.004	contradict	66.0	18.0	23.73	.000
require	97.9	82.0	12.22	.002	converge	34.0	0	24.85	.000
formulate	46.8	14.0	12.35	.002	accumulate	70.2	18.0	27.52	.000
publish	100.0	88.0	12.41	.002	investigate	83.0	30.0	28.09	.000
modify	57.4	22.0	13.42	.001	facilitate	63.8	16.0	28.44	.000
justify	74.5	40.0	14.02	.000	assess	61.7	10.0	28.82	.000
manipulate	40.4	8.0	14.10	.001	accomplish	93.6	52.0	29.71	.000
emerge	38.3	6.0	14.95	.000	evaluate	100.0	54.0	30.09	.000
attain	93.6	60.0	15.39	.000	reveal	83.0	28.0	31.50	.000
institute	27.7	2.0	15.96	.000	revise	83.0	26.0	31.65	.000
devise	40.4	8.0	16.29	.000	unify	83.0	26.0	31.65	.000
supplement	61.7	24.0	16.61	.000	refute	59.6	8.0	31.78	.000
correspond	44.7	18.0	16.69	.000	indicate	93.6	44.0	32.74	.000
specify	44.7	8.0	16.79	.000	stimulate	48.9	10.0	33.62	.000
minimise	66.0	36.0	17.30	.000	interpret	70.2	12.0	33.96	.000
underlie	70.2	28.0	17.96	.000	quote	87.2	28.0	34.83	.000
neutralize	42.6	6.0	18.61	.000	hypothesise	72.3	12.0	36.35	.000
vary	80.9	38.0	19.02	.000	shrink	95.7	38.0	37.47	.000
transform	91.5	68.0	19.36	.000	speculate	74.5	12.0	38.93	.000
participate	100.0	72.0	20.15	.000	restrict	95.7	38.0	39.24	.000
summarise	95.7	58.0	20.36	.000	inspect	87.2	30.0	40.10	.000

To understand further the differences in knowledge between the high and low groups, I calculated the cumulative total of correct answers for each verb and then ordered them from most-known to least-known. I then isolated and compared the top 30% and the bottom 30% of correct responses. The cumulative total results for the 14 most-known and 13 least-known verbs are shown in Table 4.6 below. Whereas level tests at lower levels of vocabulary, such as Nation's 1,000 and 2,000-word level tests, can be expected to produce fairly homogenous results, the findings in Table 4.6 below indicate that this pilot version of the UWL verb test produces more varied discrimination.

Table 4.6 Comparison of most-known and least-known verbs

The 14 most-known verbs			**The 13 least-known verbs**		
Verb	**Raw total of correct answers (/181)**	**% Total**	**Verb**	**Raw total of correct answers (/181)**	**% Total**
select	175	96.7	formulate	55	30.4
publish	174	96.1	refute	54	29.8
rely	170	93.9	correspond	49	27.1
affect	169	93.4	manipulate	49	27.1
require	168	92.8	specify	48	26.5
participate	160	88.4	neutralise	41	22.7
transform	156	86.2	emerge	39	21.5
fund	155	85.6	devise	38	21.0
accomplish	142	78.5	stimulate	36	19.9
evaluate	141	77.9	enumerate	35	19.3
summarise	141	77.9	function	31	17.1
attain	139	76.8	converge	25	13.8
emphasise	132	72.9	institute	22	21.2
indicate	132	72.9			

4.7.3 Correlations with overall English proficiency

The UWL verb test had a correlation of .44 with the total English proficiency score on the University of Tsukuba English Proficiency Test. For individual sections in the proficiency test, the correlations were weaker, as shown in Table 4.7. The two strongest correlations occur with vocabulary and reading, which is to be expected. However, these are relatively low correlations, which is probably due to the English proficiency test being a test of general English ability rather than of academic English alone.

Table 4.7 Correlations between academic verb knowledge and general English proficiency

1.	Proficiency full score	.44
2.	Vocabulary score	.38
3.	Reading score	.33
4.	Listening dialogues score	.32
5.	Written grammar score	.28
6.	Error recognition score	.25
7.	Listening passages score	.08

4.7.4 Insights from the think-aloud interviews

I conducted ten think-aloud interviews to investigate the processes and strategies that learners use in completing the UWL verb test. This lets us move beyond the general understanding that the VLT format measures "receptive vocabulary knowledge" (Nation, 1990) or "threshold meaning knowledge of the target words" (Schmitt, 2000) towards a finer understanding of particular types of lexical knowledge.

Each interview lasted between 40 and 60 minutes, depending on how much information each student volunteered and how quickly they completed the test. I sat to the side of each student and noted the order in which they wrote down their answers for each cluster. When each learner had completed a cluster, I asked them to explain how

they had chosen their answers in the order that they had made them. Although the interviews were not recorded, I was able to make almost verbatim notes since the students were concentrating on expressing their thoughts in English. Whenever appropriate, follow-up probing questions were used to encourage the students to articulate their thoughts more clearly.

The notes from the interviews were analyzed in four consecutive stages. The first stage involved identifying those items where the interviewees clearly responded with **know/don't know**, or described the use of partial knowledge strategies. Partial knowledge at this stage of the analysis includes both test-taking and lexical strategies reported by the interviewees. It is represented by "other" in Table 4.8 below, with the students identified by a capital letter in the top row.

Table 4.8 Use of absolute knowledge vs. partial knowledge in completing the UWL verb test

	A	B	C	D	E	F	G	H	I	J
Know	6	25	20	24	32	0	11	10	15	15
Don't know	0	6	0	5	0	9	12	0	8	14
Other	54	29	40	31	28	51	33	50	23	29
Test score /60	**47**	**38**	**41**	**43**	**40**	**29**	**40**	**41**	**47**	**34**

In the second stage of the analysis, **know** and **don't know** totals were combined to give a total for the number of items that each interviewee knew absolutely. These combined totals were re-coded as **absolute**. Items coded as **other** in the first stage were then sub-divided into two categories, **test-taking strategies** and **lexical strategies**. Test-taking strategies involve here guesses pure and simple, or a combination of knowing (but not for certain) and guessing. Lexical strategies involve exploiting several types of partial knowledge that are described in more detail in the

third stage. The second stage lets us separate global test-taking behaviour from the specific use of lexical strategies. The results are shown in Table 4.9.

Table 4.9 Absolute vs. test-taking and lexical strategies in completing the UWL verb test

	A	B	C	D	E	F	G	H	I	J
Absolute	6	31	20	29	32	9	23	10	23	29
Global test-taking strategies	13	21	15	15	15	14	18	29	4	15
Specific lexical strategies	41	8	25	16	13	37	19	21	33	16
Test score /60	**47**	**38**	**41**	**43**	**40**	**29**	**40**	**41**	**47**	**34**

In the third stage of the analysis, I sub-divided the specific lexical strategies into 10 types, following a close analysis of how the interviewees described their decision-making. These strategies are shown in Table 4.10.

Table 4.10 Specific lexical strategies in completing the UWL verb test

Strategy 1	I know what this verb means, but I can't find the exact definition here.
Strategy 2	I connect part of the definition with the verb.
Strategy 3	I make associations between the definition and the verb.
Strategy 4	I link the definition or verb to a word family.
Strategy 5	I translate the verb into Japanese and then re-translate into English.
Strategy 6	I use affixation knowledge.
Strategy 7	I paraphrase the verb and find the nearest equivalent definition.
Strategy 8	I find a connection between two words by making them overlap in meaning.
Strategy 9	I use a collocation or a phrase.
Strategy 10	I use a combination of the above lexical strategies.

Finally, I totalled use of the 10 lexical strategies for each of the interviewees. These results are displayed in Table 4.11 on the next page.

Table 4.11 **Each interviewee's use of specific lexical strategies**

Strategy	A	B	C	D	E	F	G	H	I	J	TOTAL
Strategy 1: I know what this verb means, but I can't find the exact definition here.	0	0	4	0	5	0	1	1	3	1	15
Strategy 2: I connect part of the definition with the verb.	0	1	2	1	1	0	0	1	1	1	8
Strategy 3: I make associations between the definition and the verb.	7	1	4	4	0	6	3	3	5	0	33
Strategy 4: I link the definition or verb to a word family.	2	0	0	0	1	0	4	0	1	2	10
Strategy 5: I translate the verb into Japanese and then re-translate into English.	3	0	1	0	0	5	0	0	1	0	10
Strategy 6: I use affixation knowledge.	3	1	2	2	0	0	1	0	1	1	11
Strategy 7: I paraphrase the verb and find the nearest equivalent definition.	10	2	7	4	2	21	7	10	14	4	81
Strategy 8: I find a connection between two words by making them overlap in meaning.	1	1	1	1	1	1	0	0	0	2	8
Strategy 9: I use a collocation or a phrase.	2	0	2	0	0	0	0	0	2	0	6
Strategy 10: I use a combination of the above lexical strategies.	13	2	2	4	3	4	3	6	5	5	47
TOTAL	41	8	25	16	13	37	19	21	33	16	

4.8 Discussion

I will limit the discussion to three points: (a) the differences between the high proficiency and low proficiency groups, (b) the problem of construct validity, and (c) the type of lexical knowledge exploited by learners taking this test.

4.8.1 Differences between most-known and least-known verbs

From the results in Table 4.6 above (i.e., comparison of the most-known and least-known verbs), some differences may be noted between the two groups of verbs. First, in terms of word length, one- and two-syllable verbs tend to predominate in the most-known verbs, whereas, in the least-known verbs, three- and four-syllable verbs appear to occur more frequently, as shown in Table 4.12 below.

Table 4.12 Word-length: most-known and least-known verbs

	most-known verbs	least-known verbs
1-syllable verbs	1	-
2-syllable verbs	7	5
3-syllable verbs	4	6
4-syllable verbs	2	4

This suggests that word length may be one factor in learner knowledge of such verbs, but probably a more important factor is frequency in that frequent verbs will tend to be shorter in any case. The suffix **-ise / -ize** added to nouns (here, **emphasis** and **summary**) may help the early acquisition of such verbs, but the sample is too small to draw any firm conclusions about affixation and learnability here. Finally, in semantic terms, it seems that the most-known verbs tend to display more here-and-now senses and may have fewer higher frequency near-synonyms than the least-known verbs. For

example, **select**, **rely** and **require** seem easily associable with **choose**, **depend** and **need**, whereas **formulate**, **devise** and **institute** might be less easily associated with **put together**, **plan**, and **start**. The least-known verbs, as would be expected, tend to cover much more abstract meanings.

It is difficult to judge whether the most-known verbs are more frequent than the least-known verbs, unless we use a specific academic corpus to confirm this. However, comparison with British National Corpus rankings, as calculated by Kilgariff, provides some indication that the most-known verbs (with the exception of **accomplish**, **summarise** and **attain**) fall within the top 60% of the most frequent 5,000 words of that general corpus. That is, they are located in the 3,000 most frequent BNC words. Table 4.13 below shows these frequencies, and includes Nation's own range rankings for each verb.

Nation categorizes the UWL items into 11 groups by frequency and range (Nation, 1990, pp.235-239), so that Group 1 means those words with the highest frequency and coverage, and Group 11 means those with the lowest. Just as it is not clear how Nation calculated the 11 categories, his rankings seem to be less conclusive across the two groups. On the other hand, the least-known verbs fall into the bottom 40% of the first 5,000 words of the BNC corpus, or the 3,000-5,000 most frequent range (**specify** and **emerge** excepted); furthermore, five of the least-known verbs do not even appear in the BNC corpus as measured by Kilgariff.

Table 4.13 **BNC frequencies and UWL frequencies & ranges for the most-known and least-known verbs**

The 14 most-known verbs				The 13 least-known verbs			
Verb	BNC rank	BNC frequency	Nation's frequency and range estimates	Verb	BNC rank	BNC frequency	Nation's frequency and range estimates
select	1632	5844	2	formulate	4390	1447	1
publish	835	12242	1	refute	*not listed*		7
rely	1751	54032	3	correspond	3223	2277	3
affect	789	12867	2	manipulate	4840	1244	2
require	343	28711	1	specify	2308	3760	1
participate	2785	2860	6	neutralise	*not listed*		7**
transform	2850	2762	5	emerge	1271	7706	10
fund	2362	3630	9	devise	3367	2115	1
accomplish	5648	968	3	stimulate	3457	2029	5
evaluate	3254	2238	1	enumerate	*not listed*		9
summarise	5195	1111	1*	function	3927	1698	2
attain	5016	1182	6	converge	*not listed*		10
emphasise	2748	2922	2	institute	*not listed*		5
indicate	827	12369	1				

* Listed as *summary* in Nation, 1990

** Listed as *neutral* in Nation, 1990

4.8.2 The problem of construct validity

In the explanation of the research construct and design for the present study, I noted the problem of simplification in creating the short definitions for each set of six verbs. This was the case with **select**, **rely** and **require**. At the same time, the converse is true. The less frequent the verb, the more difficult it is to create a short definition without deleting important associative semantic clues. Thus, one problem in the construct validity of the UWL verb test appears to be that the more frequent

items will tend to have easier and more understandable definitions.

A related problem with the construct validity of the test is that it is not clear whether the items are independent of each other. Here, it is clear that items both need to be constructed in sets of six words and presented to learners in such sets. It would further seem that the definitions and the distractors must necessarily be interdependently looked at in order for the test to be completed. Beglar and Hunt are somewhat contradictory on this point. Near the start of their paper, they state:

> In this study the authors have made the assumption that each item is independent even though the items appear in sets of three. Although this assumption deserves close examination and analysis, the authors chose to follow the work of previous researchers ...in assuming item independence. (Beglar & Hunt, 1999, p.138)

Towards the end, they comment:

> This study assumes that each item functions independently of the others in each set of 6 words and 3 definitions. However, it has not been shown that assumption of item independence holds true given this test format. The degree to which items in these sets interact clearly deserves further study. (Beglar & Hunt, 1999, p.154)

The retrospective protocols conducted by Schmitt, Schmitt and Clapham (2001) shed some, albeit limited, light on this question. They note:

> ... examinees tended to work down through the three stems in a serial manner. They usually reviewed all of the option words in the cluster when answering difficult items, but for easy items, they usually focused on the correct option without considering the others. (p.74)

However, this is a very general observation based on learner self-reports with just two isolated clusters. We have, then, a test that performs statistically well and that is generally accepted in the field as a reliable measure, but some very serious questions remain as to what types of lexical knowledge it tests. Here the results from the think-aloud interviews present us with some particularly interesting insights about item

independence/interdependence and the types of lexical processes and strategies that learners use.

4.8.3 Item independence and learners' lexical knowledge

The results from the 10 think-aloud interviews show that we cannot by any means assume item independence with the UWL verb test (see Tables 4.9-4.11 above). As is to be expected, the interviewees showed different totals for absolute knowledge, but the evidence across the group for guessing or using a specific lexical strategy was very clear. This supports the broad claim made by Schmitt, Schmitt and Clapham about the difference between how easy or difficult items are tackled in such a test. Table 4.9 also suggests that it is the in-between items that result in learners using various strategies. That is, when examinees do not absolutely know or not know the answer, they necessarily make use of partial knowledge, whether test-taking or lexical.

Yet, the interviewees also varied considerably in their use of lexical strategies (as shown in Table 4.11, which highlights individuals' use of specific lexical strategies). There was no clear pattern in the distribution of lexical strategies by individual, except for the use of **paraphrase** (Strategy 7). This was the most common with each student. **Associative knowledge** (Strategy 3) was the second most common individual lexical strategy, while **a combination of lexical strategies** (Strategy 10) also showed relatively strong use by each person. On the other hand, collocation or phrasal knowledge was reported by only three students in two instances in each case.

Overall, then, the apparent simplicity of the VLT format would seem to obscure the complex and variable ways in which learners exploit partial lexical knowledge. To be honest, this comes as a bit of a shock: I thought I had created a reliable instrument, but in piloting it and analyzing the results, I have gradually come to doubt whether it is

in fact suitable as a measure of lexical verb knowledge. In coming to that realization, however, I have learnt a great deal about the difficulty of creating a discrete, selective and context-independent test. So, although it somewhat surprising to realize that the VLT is rather less than perfect as a format than it at first seems, the experience of working with it has been very useful. Unexpectedly, we have gained some rather compelling insights into learners' use of partial lexical knowledge, and that is, I think, the area that I need to continue to explore in trying to find a way to measure L2 collocation knowledge outside of a textual approach.

4.9 Conclusion

In this chapter, we have investigated how to measure learners' knowledge of individual academic lexical verbs by adapting Nation's Vocabulary Levels Test. We did this for three main reasons. First, we wanted to establish a better sense of what particular academic verbs students know. Our other goals were to understand and evaluate how the VLT format works, and to see what kind of lexical knowledge it actually tests.

Through statistical analysis, we were able to establish that VLT format provides apparently reliable results, but the think-aloud interviews let us understand how learners exploit partial knowledge in taking such a test. We classified this use of partial knowledge as either test-taking behaviour (i.e., guessing without explicit recourse to lexical knowledge) or the use of specific types of lexical knowledge. In the latter case, it was difficult to see consistent patterns across the group of 10 interviewees, because individual variation was so strong. Nevertheless, the interview data showed that students tend to exploit, under these discrete, selective and context-independent conditions, three main types of lexical knowledge: **paraphrase**, **associative networks**, and **combinations of multiple lexical strategies**. That said, we need

to remember that the test format allowed us to test only 60 items in 20 minutes.

To conclude, we have moved away from a mainstream corpus-based approach towards testing vocabulary knowledge in a more experimental format. This experimentation has let us forensically dissect the VLT format so that we can see whether it is in fact suitable for a wider purpose. That purpose is to establish a de-contextualized measure of L2 collocation knowledge, and one conclusion we can draw from this study is that we need to be able to test more items in a shorter time. A second, equally important, conclusion is that we need to create a test that *explicitly* enables learners to exploit their sense of partial knowledge of L2 vocabulary. The VLT format does not let do this and so must be discarded, as it is not appropriate for the longer-term purposes that we have in mind. In the next chapter, we will consider just exactly how we can measure variable partial lexical knowledge more sensitively. That will take us closer to testing L2 collocation knowledge experimentally.

Note

1. Although Beglar and Hunt's 1999 study was the first attempt to validate the VLT, a spate of other studies of the VLT appeared within a few years: Kamimoto, 2001; Schmitt, Schmitt & Clapham, 2001; Shiotsu, 2001. Kamimoto (2001) tested Forms A and B of the VLT for the 2000, 3000, and 5000 word levels on 196 Japanese university students. He found a large number of loanwords in Japanese in the VLT; by cross-reference to BNC frequencies, he was also able to demonstrate that some items in the VLT are outdated. Schmitt, Schmitt and Clapham (2001) conducted a large-scale validation of new forms of the VLT (N = 801), as well as conducted think-aloud interviews with 22 subjects to understand both test-taking behaviour and the construct validity of the VLT. The interviewees were presented with two new clusters, one relatively easy and the other relatively difficult, and asked "retrospectively, to describe the process they had gone through as they answered items in the written test" (p.73). They were also asked to give definitions of 50 words that they had encountered in the VLT main test administration. Results from the interviews showed that the VLT seems, to a small extent, to underestimate partial lexical knowledge and to overlook guessing by test-takers. Finally, Shiotsu (2001) completed a small-scale study of two forms of the VLT (N = 40), using item response analysis to identify well-

performing and poor items. After discarding mis-fitting items, Shiotsu re-compiled items and re-combined levels of the test to create a 57-item test. Since 1999, it is clear that the VLT has become subject to greater critical evaluation, although most studies have concluded that the basic format of the test is effective, but needs improvement. A different conclusion is reached in this chapter, however, for reasons that are explained later in the discussion.

Chapter 5

Measuring States of Lexical Verb Knowledge

5.0 Overview

In the previous chapter, we evaluated the suitability of the Vocabulary Levels Test as a tool for investigating lexical verb knowledge with items drawn from the University Word List. We did this because we wanted to establish a database of lexical verbs for creating a later measure of collocation knowledge. The VLT format looked initially promising for such a purpose, and the analysis produced good item correlations as well as good reliability. However, the results from the experiment also showed that there are a number of problems with the instrument. First, only 60 items could be tested within 20 minutes, so the test was relatively inefficient for quickly producing data about a large number of items. Second, item independence was compromised because we discovered that test-takers only respond to items independently when they are certain of their answers; when they are not completely certain, they tend to work through a whole cluster to arrive at a decision for an individual item. Finally, think-aloud interviews also revealed that the VLT format encourages guessing and the use of particular types of partial lexical knowledge such as paraphrasing and associative networks. Yet, this behaviour could not be accounted for in the statistical analysis.

Overall, then, the evaluation of the VLT as a tool for measuring lexical verb knowledge let us conclude that we need to use a different instrument. An improved measure of lexical verb knowledge would ideally need to:

(i) work from a more specific construct of vocabulary knowledge;
(ii) test more items in less time;
(iii) include both high frequency and academic word list verbs;
(iv) be simpler in its format;
(v) present items independently of each other;
(vi) take explicit account of the use of partial lexical knowledge.

In this chapter, we will explore how we can respond to these needs. To do this, we will investigate how we can construct an alternative test for lexical verbs, so that the

database created can be used for a subsequent comparison of verb and noun knowledge, as well as for a later verb + noun collocation recognition study. This collocation test remains the further goal, and we will discuss a verb + noun collocation test in Chapter 6, based on the alternative format developed in this chapter.

5.1 Introduction

A useful starting point in designing an alternative instrument is to re-consider common assumptions about word knowledge. Mastery of individual words can be conventionally conceptualized in terms of the macro-categories of word form, meaning, and use (Nation, 2001). This approach to word knowledge results in a complex classification of discrete forms of word knowledge, which can be further sub-cateogorised into receptive and productive knowledge and use, as shown in Table 5.1 on the next page. This classification lets us characterize vocabulary measures by the specific types of receptive or productive word knowledge that they aim at testing.

Apart from the receptive-productive distinction, there are two further major dimensions by which vocabulary measures can be conventionally distinguished—breadth of vocabulary knowledge and depth of vocabulary knowledge. The VLT is usually referred to as a receptive measure of vocabulary size or breadth, while a test requiring the composition of sentences to illustrate the meaning of different words would be seen as a productive measure of vocabulary depth. While these pairs of descriptors—**receptive/productive, breadth/depth**—are useful at a general level of description and classification, they are nevertheless idealisations. In the case of the VLT, for example, the findings from the UWL verb test think-alouds showed that partial knowledge is a critical factor in test performance, i.e., learners drew partially on their depth of knowledge in order to decide their responses.

Partial lexical knowledge cuts across the dimensions of breadth and depth and also raises questions as to whether a clear-cut distinction can be maintained between

Table 5.1 Discrete types of word knowledge (Nation, 2001, p.347)

Form	spoken	Receptive	Can the learner recognize the spoken form of the word?
		Productive	Can the learner produce pronounce the word correctly?
	written	Receptive	Can the learner recognize the written form of the word?
		Productive	Can the learner produce spell and write the word?
	word parts	Receptive	Can the learner recognize known parts in the word?
		Productive	Can the learner produce appropriate inflected and derived forms of the word?
Meaning	form and meaning	Receptive	Can the learner recall the appropriate meaning of this word form?
		Productive	Can the learner produce the appropriate word form to express this meaning?
	concept and referents	Receptive	Can the learner understand a range of uses of the word and its central concept?
		Productive	Can the learner use the word to refer to a range of items?
	associations	Receptive	Can the learner produce common associations for this word?
		Productive	Can the learner recall this word when presented with related ideas?
Use	grammatical functions	Receptive	Can the learner recognize correct uses of the word in context?
		Productive	Can the learner use this word in the correct grammatical patterns
	collocations	Receptive	Can the learner recognize appropriate collocations?
		Productive	Can the learner produce the word with appropriate collocations?
	constraints on use	Receptive	Can the learner tell if the word is common, formal, infrequent, etc.?
	(register, frequency...)	Productive	Can the learner use the word at appropriate times?

receptive and productive lexical knowledge.

In this chapter, I will pursue these issues further. I will start by re-considering the complex partial knowledge decision-making that learners may engage in. Second, I will critically review other instruments for estimating different degrees of lexical knowledge. This will allow us to identify specific practical and theoretical problems which need to be addressed in designing an alternative instrument for testing lexical verb knowledge. I will then present the design and development of such an instrument and analyse its performance.

5.2 Interpreting complex partial knowledge decision-making

In this section, I will take a closer look at interpreting complex partial knowledge decision-making so that we can determine what kind of instrument is appropriate for measuring such knowledge.

5.2.1 Complex partial knowledge decision-making

In the UWL verb test evaluated in Chapter 4, learners were required to match a short definition of a word for a particular word form to demonstrate minimum "receptive knowledge" of word meaning. However, even with this apparently simple matching definition format, we saw that learners use other knowledge and do not always recognise a known definition for a particular word. A brief review of one learner's think-aloud with two sets of items of the UWL verb test illustrates the lexical decision-making involved. In both sets, the learner makes completely correct responses.

The first set (Items 4-6) shows the learner making decisions very much in the way that the test claims to work.

4. judge the worth of	**a. withdraw**	**4. d**
	b. dissolve	**5. f**
5. print a book or magazine	**c. concentrate**	**6. e**
	d. evaluate	
6. take in water or knowledge	**e. absorb**	
	f. publish	

The learner's think-aloud went as follows:

> "*Print a book or magazine* reminds of publisher, so I chose **f. publish**. *Judge* equals evaluate something, so I decided **d. evaluate**. When I look at *take in*

> *water or knowledge*, I see *water*, and *absorb* means *sponge, absorbing water*, so I choose **e. absorb**."

For each item, the learner is able to match the short definition with one of the possible responses, without cross-referencing to any of the distractors. In contrast, with the second set, the same learner draws on different types of word knowledge to arrive at three correct answers. The second set involves Items 19-21:

19. decide on the meaning of	**a. persist**	**19. d**
	b. shrink	**20. f**
20. arrange in a systematic way	**c. survey**	**21. b**
	d. interpret	
21. become smaller	**e. generate**	
	f. structure	

This time, the same learner reported:

> "I know the meaning of *interpret*, so I choose **d. interpret** for *decide on the meaning of*. I cannot imagine *systematic way*, so I go to *become smaller*. I'm not sure what **a. persist** means, and **e. generate** means *become greater*, so I deleted this one. **b. shrink** makes me imagine the movie Shrink or Shrek and small animals, so I choose **b. shrink** for 21. **f. structure** makes me think of mechanical or building, so *systematic* will be mechanical or ordered, so I choose **f.**"

Again, the learner has three correct responses, but we can see clearly that her guessing of both **shrink** and **structure** shows the use of partial word knowledge. If this is characteristic of how learners use such knowledge, we need to consider how we can narrow down the concept so that a simple and reliable test instrument can be created.

5.2.2 The vocabulary scale

Vocabulary scale studies are worth considering because, similar to the VLT, such instruments have gained widespread, but somewhat uncritical, popularity among teachers and researchers interested in measuring partial knowledge. A vocabulary scale requires learners to self-report and, when necessary, demonstrate how well they know the different words in a particular test. Here, I will briefly review different types of vocabulary scale that have been used.

Working with L1 subjects in the UK, Heim and Watts (1961) used a self-judging vocabulary scale with tests comprising 80 words. In the first half of the test, the subjects were required to make one of three judgments for 40 words:

A: **I know the word and can explain the meaning to someone unfamiliar with it.**
B: **I am doubtful.**
C: **I have absolutely no idea.** (Heim & Watts, 1961, p.176)

In the second part of the test, the subjects were presented with the same 40 words again, but this time each individual word was followed by six phrases from which they had to choose the one closest in meaning to the target word. Dale[1] (1965) suggested that the following 4-point scale could be used to test American children's vocabularies:

Stage 1: **I never saw the word before.**
Stage 2: **I know there is such a word, but I don't know what it means.**
Stage 3: **I recognize the word, but I don't know specifically what the word means.**
Stage 4: **I know it.** (Dale, 1965, p.898)

With Stage 3, the learners need to supply another word that they know, so Dale's scale includes production as well as claims of strength of recognition.

Wesche and Paribakht (1996; Paribakht & Wesche, 1997) employed a 5-stage self-reporting instrument (**The Vocabulary Knowledge Scale**) that similarly includes recognition and production for several types of word knowledge:

Self-report category 1: **I don't remember having seen this word before.**
Self-report category 2: **I have seen this word before, but I don't know what it means.**
Self-report category 3: **I have seen this word before, and I *think* it means _____ . (synonym or translation)**
Self-report category 4: **I *know* this word. It means _____ . (synonym or translation)**
Self-report category 5: **I can use this word in a sentence: _____ .**
(If you do this section, please also do Section IV.) (Wesche & Paribakht, 1996, p.30)

W&P used the scale to track the development of learners' vocabulary knowledge in a content-based ESL reading course. Their 1996 paper reports results for 32 words and states that the VKS "could capture progression in the development of knowledge of particular words" (Wesche & Paribakht, 1996, p.30). Tracking these changes through pre- and post-tests, the authors used a differential scoring system where 1 point was

potentially given for Category 1, 2 points for Category 2, 3 for Category 3, 4 for Category 4, and 5 points for Category 5. The first two categories involved self-report only, while the other three categories required learners to demonstrate word knowledge. Wrong responses for Categories 3-5 were given a score of 2. In both of their studies, learners were found to self-report their knowledge reliably. The VKS scoring system for their 1996 study is shown in Figure 5.1 below.

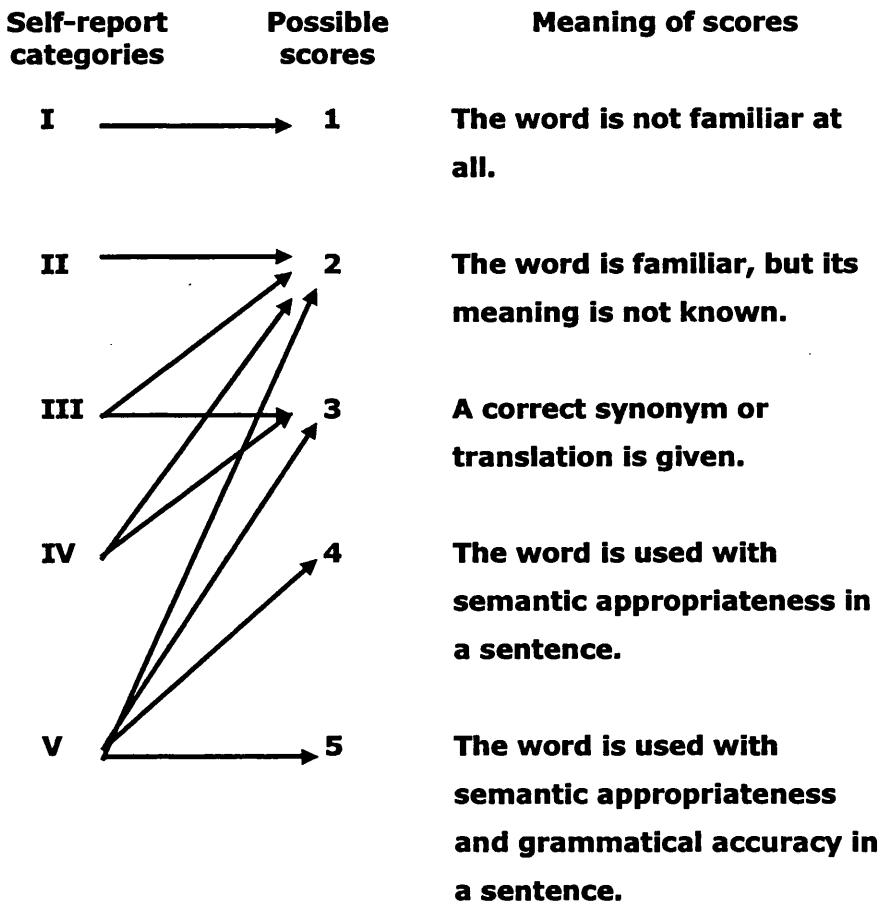

Figure 5.1 **VKS scoring categories (Wesche & Paribakht, 1996, p.30)**

Joe (1995) reports on a study targeting 12 low-frequency words, where she adapted the VKS into 6 stages as the "demonstrated word knowledge scale" (Joe, 1995, p.151). Joe's scale consisted of the following six categories:

Score	Interpretation
1	The word is not familiar at all.
2	The word is familiar but the meaning is unknown. An affix is familiar but the base and general meaning is unknown.

> **3 The word is familiar. An association or general meaning is known.**
> **4 The word's specific meaning is known. The word is used accurately and appropriately in a sentence, but only the general meaning is supplied.**
> **5 The word is used with semantic appropriateness in a sentence.**
> **6 The word is used with semantic appropriateness and grammatical accuracy in a sentence (also uses acceptable collocations).** (Joe, 1995, p.151)

In contrast, Scarcella and Zimmerman (1998) used a simplified VKS called the Test of Academic Lexicon (TAL) with 192 university ESL students. The test consisted of 40 real words and 10 non-words, and the students were asked to indicate their knowledge on one of four levels:

> **(a) I don't know the word.**
> **(b) I have seen the word before but am not sure of the meaning.**
> **(c) I understand the word when I see or hear it in a sentence, but I don't know how to use it in my own speaking or writing.**
> **(d) I can use the word in a sentence.** (Scarcella & Zimmerman, 1998, p.36)

If the (d) was chosen, the students were asked to compose an example sentence showing how the word is used. A differential scoring was also used in this study: 4 points for correctly answering (d), 3 points for (c), 2 for (b), and 1 for (a), and 0 points for an incorrect answer for (d).

In sum, a substantial amount of work has been done with vocabulary scales, and strong claims have been made about the effectiveness of such scales for measuring word knowledge. What, though, are the weaknesses with such a measure?

5.2.3 Weaknesses of the vocabulary scale

I would like to focus on Wesche and Paribakht's VKS to highlight a number of problems with trying to use a "scale" to measure word knowledge. One of the major weaknesses of the VKS is that scores tend to cluster towards 2 from different categories ("I have seen this word before, but I don't know what it means."). It is neither clear why wrong responses in Categories 3-5 should be re-classified as Category 2, nor obvious why some responses in Category 4 might be scored as Category 3, and wrong

responses in Category 5 as either Category 3 or 4. The difference between "semantic appropriateness" and "semantic appropriateness and grammatical accuracy" remains, for example, uncertain. Another weakness is that the five categories mix different word knowledge criteria. A further problem lies in unpacking the scores from the test. Suppose two learners, A and B, have scored 20 on a 10-item VKS test. Their scores could have been produced by many different combinations of 1-5, for example:

Table 5.2 The problem of interpreting VKS scores

Item	1	2	3	4	5	6	7	8	9	10
A	1	5	1	3	1	4	1	1	2	1
B	4	1	2	1	2	1	2	2	3	2

Both learners have a mean score of 2, yet show very different profiles. The VKS is also very time-consuming to check and score, so the instrument can only be used with a small number of target items and learners. In short, the instrument is cumbersome, complicated, and oblique.

A more serious issue concerns the theoretical assumption of a continuum with discrete stages from passive to active knowledge through which words pass as they become more established in a second language user's mental lexicon. Although W&P do not intend to "claim that vocabulary acquisition is essentially linear" (Wesche & Paribakht, 1996, p.29), it is difficult to separate the notion of a scale from that of linearity, just as it is hard not to interpret the different stages as in some kind of linear relationship: "This instrument captures in a relatively efficient way certain stages in the initial development of core knowledge of given words" (Wesche & Paribakht, 1996, p.29) is how W&P also describe the scale, for example. Melka Teichroew (1982, p.21) succinctly describes the characteristics of linearity in these terms:

> The notion of familiarity (or knowledge) of a word could be represented as being a line, a continuum starting, roughly, with the first stage of recognition, passing through various intermediary points and finishing near productive knowledge, with productive knowledge itself composed of several stages or phases...

Similarly, the previous analysis of the VKS suggests that "familiarity" leads into "synonymy," which in turn leads into "semantic appropriacy." In short, (at least) three distinct types of partial lexical knowledge are interrelated to create a single contiguous "scale."

To conclude, the VKS uses, as do other scales, multiple types of word knowledge. Instead of narrowing down the concept of partial knowledge, the proliferation of categories broadens the construct. This leaves the instrument theoretically weak. W&P state that "the VKS should be viewed as a practical instrument for use in studies of the initial recognition and use of new words" (1996, p.29). In as much as the categories in the VKS modestly reflect "naturalistic" stages of learning, the VKS is perhaps closer to a pedagogic device than a research tool. Having noted several problems with the VKS and its derivatives, I will now consider an alternative instrument.

5.2.4 An alternative to the vocabulary scale

A useful step in limiting the concept of partial knowledge is to focus on "knowledge" as the common dimension between different components of the instrument. Rather than select specific sub-types of lexical knowledge for a test instrument, we might consider different states of knowledge as a single dimension of lexical knowledge. Schmitt (2000, p.167) suggests as much in noting:

> ...vocabulary learning is incremental, and so mastery of these aspects will tend to vary on a continuum stretching from "no knowledge" at one end to "complete knowledge and control" at the other.

Yet, the continuum metaphor is misleading in implying incremental increases through sequential stages from "nothing" to "complete."

A different line of thinking is taken by Meara (1999) who argues that we can better understand learners as knowing lexical items in different states such as "not known" and "automatically known." According to Meara, any word can move from one state to any of the other knowledge states during a particular period of time. For example, a word can move from "automatically known" to "not known," or "known, but have to think about it" and so on. At the same time, if we score each item as 1 for whatever state it is placed in, we have a much simpler and more transparent method

for calculating individual performance. The four knowledge states used in Meara's model are represented in Figure 5.2 below.

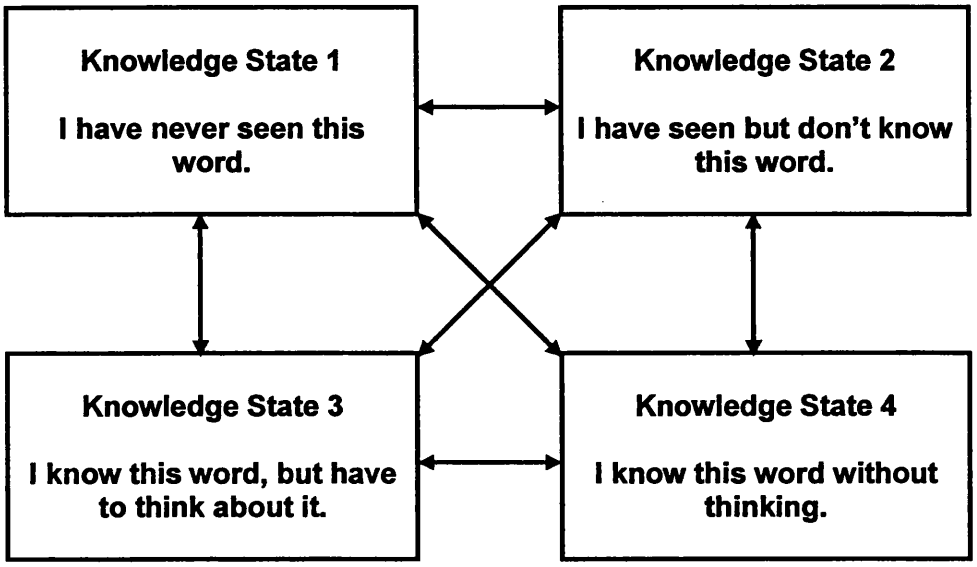

Figure 5.2 Four states of lexical knowledge (based on Meara, 1999)

Although I have characterized this model as using a single dimension of lexical knowledge, this is not quite the case. The single dimension has two aspects, **recognition** (*see / know*) and **automaticity** (*have to think about it / without thinking*). Yet, it is the variable combination of these two aspects that allows the four states to be seen as separate. The descriptors also have a strong psychological reality for learners, so they are easy to use. Waring (2002) proposes a similar model along the dimension of how well learners know words:

A. I do not know this word.
B. I know this word a little.
C. I know this word quite well.
D. I know this word very well.

However, differentiating *a little*, *quite well*, and *very well* has, in my experience, proved to be more difficult for learners than responding to the descriptors given in Figure 5.2 above.

Overall, then, a knowledge state model looks practically and theoretically sounder than a scale model. While I earlier noted several problems with vocabulary scales, fewer weaknesses were identified with the knowledge state ranking instrument. This evaluation provides us with good reasons for believing that the state model has

greater potential and is more appropriate. In the study presented in the rest of this chapter, I report on the development of a new lexical verb test that exploits knowledge states. We will look at the performance of the test and analyse the results in order to present a rationale for selecting verbs for a following verb + noun collocation experiment.

5.3 Research goals

The goals of the experiment are to examine:

1) how well the alternative instrument works;
2) which lexical verbs students know;
3) whether lexical verb knowledge is related to lexical verb frequency.

5.4 Research method

The verb recognition test is the first of three interrelated experiments. The second is a noun recognition test, and the third a verb + noun collocation recognition test, both of which are presented in the next chapter. By conducting the present experiment with an increased number of lexical verbs, we will have a large pool of items to choose from for the later verb + noun collocation recognition test. Moreover, the three separate measures (verb recognition, noun recognition, and verb + noun collocation) will allow us to judge how knowledge of individual lexical items (verbs and nouns) is connected to knowledge of lexical combinations (verb + noun collocations).

In the verb recognition test, 150 lexical verbs were targeted for knowledge state ranking with the following instrument:

1 I have never seen this verb.
2 I have seen but don't know this verb.
3 I know this verb but have to think about it.
4 I know this verb without thinking.

The four states of knowledge are adapted from Shillaw's parallel work into learner recognition of nouns (Shillaw, 2001). Knowledge state ranking means here degree of learner recognition knowledge.

The instrument was chosen for the following reasons. First, the four states directly represent partial lexical knowledge, but do not sub-categorise the construct beyond the two general aspects of recognition and automaticity. The instrument is therefore simpler than either the VLT or the VKS. Second, the instrument asks

learners to self-report their knowledge in one of four clearly distinguishable categories, but it does require production of knowledge. This means that the measure is one of recognition alone, and that overlaps between recognition and production are avoided. The third reason is that the instrument allows each response to be scored as 1 or 0. Scoring can be simplified and kept consistent from item to item and subject to subject. Next, the instrument treats each item separately without reference to any other items or distractors, so item independence is not open to question. Finally, a much greater number of verbs can be tested in a shorter time (150 verbs in 10-15 minutes, compared to 60 verbs in 20-25 minutes with the UWL verb test), which means that a wider selection of lexical verbs can be included in the measure. For these reasons, the revised test instrument meets the six areas for improvement presented in the opening to this chapter.

Various factors influenced the selection of verbs for this test. The results from the UWL verb test showed that few items were known by more than 70% of the population. That test may have in fact been too difficult, so I felt it was wise to include not only "academic" verbs for the present experiment, but also higher frequency "core" verbs. I took this decision in order to ensure that a much greater number of items would in theory be confidently known, which in turn would create a more solid item bank for choosing verb + noun collocations for the following experiment. For the current experiment, 50 of the items were selected from the first 2000 words of the General Service List (West, 1953): 25 from the first 1000 range, and 25 from the 1000 to 2000 range. The other 100 verbs chosen for this experiment came from the first 4 ranges of the Academic Word List (Coxhead, 2000). This was done to allow for a greater number of the most widely used academic verbs to be tested. (It is perhaps important to note that the AWL is linked to the GSL for its selection criteria, a point that I will return to later when I examine the results of the test; for the time being, in terms of construct validity, we may simply note that the GSL and AWL are closely related.) The selection of verbs for the whole test yielded a final total of 149 verbs, since one verb was inadvertently included twice in the final verb selection. The verbs are shown in Table 5.3 below, arranged in alphabetical order within each of the six ranges.

Table 5.3 Verbs selected for knowledge-state ranking

GSL 1-1000	GSL 1000-2000	AWL Range 1	AWL Range 2	AWL Range 3	AWL Range 4
admit	approve	analyse	achieve	comment	code
adopt	avoid	approach	acquire	compensate	commit
answer	borrow	assess	affect	consent	communicate
ask	calculate	assist	compute	constrain	concentrate
break	collect	assume	conclude	contribute	confer
build	compose	attribute	conduct	convene	contrast
change	discuss	consist	construct	coordinate	debate
consider	dismiss	constitute	consume	correspond	emerge
count	explore	create	credit	deduce	implement
declare	govern	define	distribute	dominate	implicate
defend	hurt	demonstrate	evaluate	ensure	impose
develop	interrupt	derive	focus	exclude	integrate
employ	manage	establish	injure	illustrate	investigate
explain	obey	estimate	institute	immigrate	occupy
feed	postpone	export	invest	imply	predict
forget	preserve	finance	maintain	interact	project
include	produce	function	obtain	justify	promote
mention	recommend	identify	participate	link	retain
observe	rob	indicate	perceive	locate	stress
possess	scratch	involve	purchase	maximise	undertake
protect	solve	legislate	regulate	negate	
refuse	spoil	occur	restrict	publish	
seize	threaten	proceed	seek	react	
trust	whisper	require	select	register	
write		research	survey	rely	
		respond	transfer	remove	
		vary		specify	

The knowledge state ranking instrument and test instructions were translated into Japanese, and the test was administered to 217 students. The students belonged to 7 different class groups: Area Studies (17), Engineering Systems A (27), Environmental Science A (37), Humanities A (36), Medical Science A (30), Physical Education A (36) and Social Management and Policy A (34). All of the students were first-year undergraduates except for the Area Studies group, a postgraduate class.

In terms of proficiency levels, as judged by the University of Tsukuba English placement test and proficiency exam, the students ranged from low-intermediate to high-intermediate. The three lowest proficiency groups were Engineering Systems, Physical Education and Environmental Science. Social Management and Policy and Humanities ranged above those groups within a mid-intermediate band, whereas Medical Science ranked as high-intermediate. As for the Area Studies group, no common objective test scores were available for assessing their English proficiency, although in terms of reading and writing ability they varied from mid to high intermediate.

5.5 Results

The data for the verb recognition test were analysed in three ways. For subject-based analysis, the results for each knowledge state were separately scored. For each knowledge state, learner claims of recognition were scored as 1, and descriptive statistics were calculated. This allows us to see how the instrument generally performs across the four states of knowledge. An analysis of variance was then carried out to see whether there were significant differences between the four knowledge states and whether the learners scored significantly better on any particular knowledge state. Correlations between learners' English proficiency and verb recognition were also analysed. The 149 verbs were then ordered by the totals for Knowledge State 4 for each verb, so that a rank order listing of the verbs could be produced. The rank order was also cross-referenced to the six source ranges (see Table 5.3 above) and to British National Corpus frequency rankings (Kilgarriff, 1996) to determine whether there is a relationship between degree of learner recognition and frequency. We will continue by looking at each set of results in more detail.

5.5.1 Results: descriptive statistics

The results for each Knowledge State are shown in Table 5.4 below. The mean for Knowledge States 1, 2, and 3 is consistently low (6.5, 19.86, and 18.49, respectively), while it is much higher for Knowledge State 4 (97.92). Conversely, the s.d. values for Knowledge States 1, 2, and 3 are relatively high, while the s.d value for Knowledge State 4 (22.12) is comparatively much lower.

Table 5.4 **Lexical verbs assigned to each knowledge state**

	KS1	KS2	KS3	KS4
N	217	217	217	217
Mean	6.5	19.86	18.49	97.92
Standard deviation	6.31	13.24	13.88	22.12
Cronbach's α	.88	.92	.93	.97

The reliability of the instrument, as measured by Cronbach's alpha, is very high for Knowledge States 2, 3, and 4, but drops for Knowledge State 1 (.88) owing to zero variance with a number of items (although .88 is still a very solid result).

5.5.2 Results: comparison of subjects for the four knowledge states

Each student's total score was calculated for each knowledge state to produce four separate scores for each student. To see whether the knowledge states generated significantly different scores by subject, I ran an ANOVA on those four sets of total scores. The results are shown in Table 5.5 below.

Table 5.5 Comparison of total scores for the four knowledge states (*p* values)

	KS1	KS2	KS3	KS4
KS1	-	.001	.737	.001
KS2	.001	-	.670	.001
KS3	.737	.670	-	.001
KS4	.001	.001	.001	-

Students scored significantly better on Knowledge State 4 compared to the other three knowledge states. They also showed significant differences between Knowledge States 1 and 2. Knowledge State 3 showed the weakest relationship to the other three states in that it was significant only to Knowledge State 4.

5.5.3 Results: comparison of different groups

I carried out a further subject-based analysis by comparing the seven different class groups. A one-way ANOVA test for Knowledge State 4 pointed to significant variation between groups: $F(6, 210)=10.173$, $p<.001$). This was confirmed by a post-hoc Scheffe test. Significant difference was found between Medical Science (high intermediate), on the one hand, and two low intermediate groups, Engineering Systems and Physical Education, on the other. A second one-way ANOVA test for Knowledge States 3 & 4 combined was carried out to see if other significant differences between groups could be found for lexical verbs "claimed as more or less known". The results were: $F(6, 210)= 8.451$, $p<.001$). As with the first ANOVA, significant difference was located only between Medical Science, on the one hand, and Engineering Systems and Physical Education, on the other.

When the groups were placed by mean group scores, they followed the same order as their general English proficiency level. This is shown below in Figure 5.3 for Knowledge State 4 and Figure 5.4 for Knowledge States 3 & 4 combined.

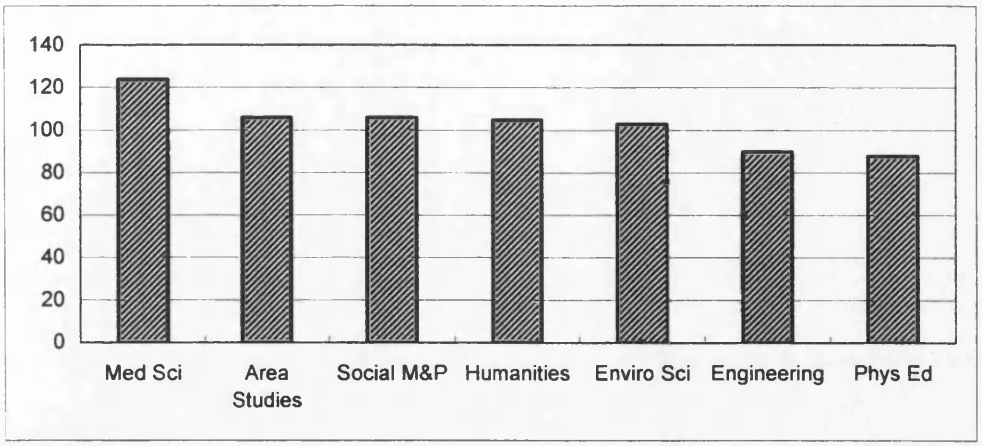

Figure 5.3 Group mean scores for Knowledge State 4

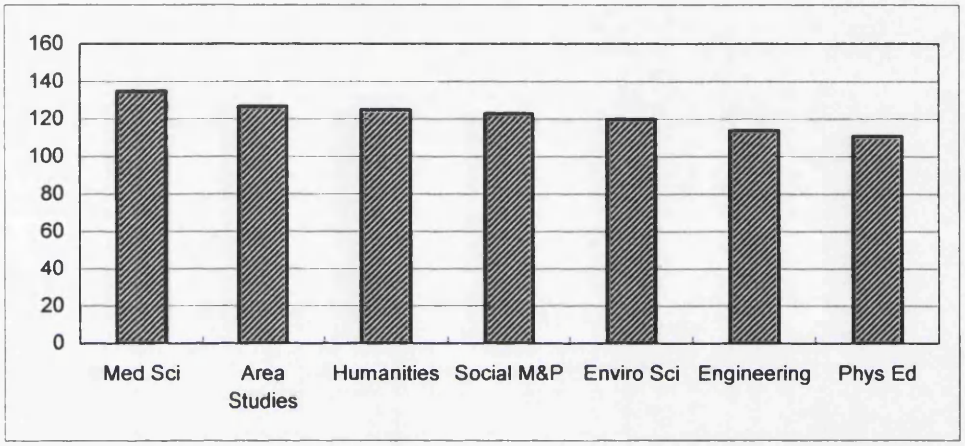

Figure 5.4 Group mean scores for Knowledge States 3 & 4 combined

In terms of subject-based analysis, the comparison of group results indicates that the verb recognition test produces significant correlations with broad bands of proficiency English (here, high intermediate and low intermediate), but cannot distinguish significantly between groups in the middle of the intermediate band. This was confirmed by a Scheffe test that identified significant differences between one high group (Medical Science) and two low groups (Engineering Systems and Physical Education), only. Although two middle groups, Humanities and Social Management & Policy, overlapped with both higher groups and lower groups, the Scheffe test did not indicate a significant difference between the three levels of proficiency of high, mid, and low as subsets, $p < .05$. This is probably because there is some overlap in

proficiency between all groups, and the overlap is strongest with learners of mid-intermediate proficiency.

5.5.4 Results: rank order of lexical verbs

The rank order results for the verbs were organized by the results for Knowledge State 4. Table 5.6 below give these results and includes the separate totals for each knowledge state, as well as the combined totals for Knowledge States 1 + 2 (Column A) and 3 + 4 (Column B). Columns A and B have been included to present a broad indication between the binary distinction of "claimed as more or less unknown" and "claimed as more or less known." The last two columns show the source frequency range and the British National Corpus rank for each verb.

Table 5.6 **Verbs in rank order for Knowledge State 4, with source frequency range and BNC rank (N=217)**

Verb	State 1	State 2	State 3	State 4	Total States: 1 + 2 [A]	Total: States 3 + 4 [B]	Source frequency range	BNC rank
ask	0	0	0	217	0	217	GSL1	154
change	0	0	0	217	0	217	GSL1	351
answer	0	0	0	217	0	217	GSL1	989
develop	0	0	1	216	0	217	GSL1	411
write	0	1	0	216	1	216	GSL1	223
forget	1	0	1	215	1	217	GSL1	828
build	1	0	1	215	1	216	GSL1	418
discuss	1	2	2	214	3	216	GSL2	690
research	0	1	2	214	1	216	AWL1	4009
communicate	0	0	4	213	0	217	AWL4	2933
explain	1	1	2	213	2	215	GSL1	537
create	2	1	1	213	3	214	AWL1	482
break	0	3	1	213	3	214	GSL1	532
select	0	2	3	212	2	215	AWL2	1632
collect	0	0	6	211	0	217	GSL2	1298
produce	0	1	5	211	1	216	GSL2	332
approach	0	0	7	210	0	217	AWL1	1310
protect	0	1	6	210	1	216	GSL1	1174
solve	0	2	6	209	2	215	GSL2	2262
employ	1	1	7	208	2	215	GSL1	1244
manage	0	1	10	206	1	216	GSL2	688
publish	0	2	9	206	2	215	AWL3	835
count	1	2	10	204	3	214	GSL1	1923
borrow	2	5	6	204	7	210	GSL2	2499

focus	2	2	11	202	4	213	AWL2	1848
refuse	0	2	14	201	2	215	GSL1	979
concentrate	0	5	11	201	5	212	AWL4	1355
consider	0	2	15	200	2	215	GSL1	349
remove	0	5	13	199	5	212	AWL3	891
hurt	1	4	13	199	5	212	GSL2	1884
avoid	0	5	14	198	5	212	GSL2	866
maintain	0	5	15	197	5	212	AWL2	856
project	0	5	16	196	5	212	AWL4	2872
trust	1	9	11	196	10	207	GSL1	2130
establish	0	5	18	194	5	212	AWL1	596
mention	1	6	16	194	7	210	GSL1	921
comment	0	6	18	193	6	211	AWL3	1837
whisper	1	13	10	193	14	203	GSL2	2613
obey	0	11	14	192	11	206	GSL2	4685
debate	2	12	11	192	14	203	AWL4	3585
achieve	0	5	21	191	5	212	AWL2	611
identify	1	8	17	191	9	208	AWL1	797
occur	1	4	23	189	5	212	AWL1	650
conclude	0	9	19	189	9	208	AWL2	1764
stress	0	7	22	188	7	210	AWL4	2204
require	0	8	23	186	8	209	AWL1	343
feed	1	9	22	185	10	207	GSL1	2286
include	0	12	20	185	12	205	GSL1	272
function	2	11	19	185	13	204	AWL1	3927
defend	0	11	22	184	11	206	GSL1	2182
demonstrate	3	5	27	182	8	209	AWL1	1507
explore	3	10	22	182	13	204	GSL2	1979
seek	3	13	19	182	16	201	AWL2	624
assist	1	12	23	181	13	204	AWL1	2202
justify	3	11	22	181	14	203	AWL3	2205
export	3	13	20	181	16	201	AWL1	3728
link	3	14	20	180	17	200	AWL3	1285
locate	0	16	22	179	16	201	AWL3	2323
recommend	1	12	26	178	13	204	GSL2	1618
contrast	1	13	25	178	14	203	AWL4	4589
observe	0	13	28	176	13	204	GSL1	1323
injure	5	16	20	176	21	196	AWL2	2945
promote	0	15	28	174	15	202	AWL4	1571
respond	0	18	25	174	18	199	AWL1	1348
involve	1	16	27	173	17	200	AWL1	341
analyse	12	17	16	172	29	188	AWL1	2166
calculate	3	22	21	171	25	192	GSL2	2246
react	6	21	20	170	27	190	AWL3	2966
illustrate	5	13	30	169	18	199	AWL3	1822
occupy	6	14	28	169	20	197	AWL4	2021
admit	0	9	40	168	9	208	GLS1	925
rely	6	21	22	168	27	190	AWL3	1751
transfer	2	26	21	168	28	189	AWL2	1612
interrupt	2	23	25	167	25	192	GSL2	3354

threaten	15	15	23	164	30	189	GSL2	1475
define	1	28	25	163	29	188	AWL1	1097
possess	8	22	25	162	30	187	GSL1	2194
consist	0	15	41	161	15	202	AWL1	1655
spoil	5	29	23	160	34	183	GSL2	4373
affect	0	10	47	158	10	205	AWL2	789
adopt	2	10	47	158	12	205	GSL1	1186
participate	1	26	32	158	27	190	AWL2	2785
immigrate	4	22	34	157	26	191	AWL3	0
acquire	6	18	37	156	24	193	AWL2	1512
vary	7	26	28	156	33	184	AWL1	1401
credit	1	25	38	153	26	191	AWL2	5421
survey	2	28	34	153	30	188	AWL2	4572
postpone	9	31	25	152	40	177	GSL2	5140
govern	8	34	26	149	42	175	GSL2	2938
scratch	10	33	25	149	43	174	GSL2	5056
rob	18	33	19	147	51	166	GSL2	5631
obtain	1	34	36	146	35	182	AWL2	824
indicate	1	32	48	136	33	184	AWL1	827
compose	2	41	38	136	43	174	GSL2	3548
ensure	7	40	35	135	47	170	AWL3	736
conduct	2	28	55	132	30	187	AWL2	1835
contribute	1	30	54	132	31	186	AWL3	1553
estimate	6	46	37	128	52	165	AWL1	1834
commit	3	24	64	126	27	190	AWL4	1426
emerge	14	43	34	126	57	160	AWL4	1271
finance	4	40	48	125	44	169	AWL1	2360
purchase	16	44	32	125	60	157	AWL2	2305
investigate	7	55	31	124	62	155	AWL4	1804
declare	8	51	37	121	59	158	GSL1	1583
construct	4	51	43	119	55	162	AWL2	2225
specify	3	62	34	118	65	152	AWL3	2308
dominate	11	51	38	117	62	155	AWL3	2129
predict	7	59	34	117	66	151	AWL4	2396
evaluate	23	43	34	117	66	151	AWL2	3254
coordinate	29	41	30	117	70	147	AWL3	3703
consume	6	55	41	115	61	156	AWL2	4841
imply	12	54	38	113	66	151	AWL3	1731
approve	1	44	62	110	45	172	GSL2	1780
maximise	42	32	34	109	74	143	AWL3	0
seize	27	55	27	108	82	135	GSL1	3002
code	11	61	39	106	72	145	AWL4	4582
proceed	3	62	48	104	65	152	AWL1	2201
dismiss	13	60	40	104	73	144	GSL2	2148
assume	3	45	66	103	48	169	AWL1	912
exclude	8	50	56	103	58	159	AWL3	2077
regulate	13	52	50	102	65	152	AWL2	4055
perceive	5	55	57	100	60	157	AWL2	2815
register	11	58	48	100	69	148	AWL3	2140
invest	15	69	33	100	84	133	AWL2	2384

correspond	5	57	57	98	62	155	AWL3	3223
undertake	16	66	38	97	82	135	AWL4	1695
preserve	5	50	66	96	55	162	GSL2	2363
consent	9	66	49	93	75	142	AWL3	0
constitute	8	57	61	91	65	152	AWL1	2253
interact	19	60	47	91	79	138	AWL3	0
institute	0	54	73	90	54	163	AWL2	0
compute	20	72	40	85	92	125	AWL2	6044
restrict	16	84	41	76	100	117	AWL2	2193
legislate	43	58	41	75	101	116	AWL1	0
distribute	5	70	69	73	75	142	AWL2	2787
derive	54	60	36	67	114	103	AWL1	1836
attribute	12	76	64	65	88	129	AWL1	3187
impose	28	80	51	58	108	109	AWL4	1517
compensate	39	93	37	48	132	85	AWL3	4603
retain	28	98	46	45	126	91	AWL4	1547
integrate	35	105	34	43	140	77	AWL4	3202
assess	55	85	40	37	140	77	AWL1	1570
implicate	39	108	39	31	147	70	AWL4	0
deduce	83	83	20	31	166	51	AWL3	2182
confer	55	86	51	25	141	76	AWL4	5021
convene	78	90	30	19	168	49	AWL3	0
negate	119	55	24	19	174	43	AWL3	0
implement	77	97	32	11	174	43	AWL4	2243
constrain	81	106	20	10	187	30	AWL3	5875

The rank ordering of the lexical verbs shows an even distribution of scores from most known at the top of the table to least known at the bottom. By looking down Columns A and B, it becomes clear that roughly 90% of the verbs are claimed as more or less known by more than half of the students. The verb **attribute** marks the border where the majority of students start to recognise particular verbs as more or less unknown. There are 14 verbs in this group (**derive, attribute, impose, compensate, retain, integrate, assess, implicate, deduce, confer, convene, negate, implement,** and **constrain**), and they come exclusively from AWL ranges. On the other hand, we can see at the top of the table that just a handful of the most known verbs come consistently from the GSL1 source—the most frequent verbs in English.

5.6 Discussion

We will limit the discussion to considering (a) how well the test instrument works, (b) patterns in learner recognition of lexical verbs, and (c) the relationship between learner recognition and verb frequency.

5.6.1 Evaluating the test instrument

The lexical verb recognition test involved the piloting of an alternative instrument, namely a knowledge state ranking. The four states of knowledge were:

1 **I have never seen this verb.**
2 **I have seen but don't know this verb.**
3 **I know this verb but have to think about it.**
4 **I know this verb without thinking.**

The analysis of variance by subject for their total scores in each knowledge state (see Table 5.5 above) shows that learners make a significant difference between Knowledge State 4 and the other states. With this set of 149 verbs, learners are claiming certain and automatic knowledge of just under two thirds of the target items. This is shown in Table 5.4 by the mean of 97.92 for Knowledge State 4. It seems likely that many of these automatically recognized verbs will be from the most frequent source ranges, and this is a point that I will address later in the discussion in 5.6.3 under the rubric of 'The relationship between learner recognition and verb frequency'. Table 5.5 shows significant differences between Knowledge State 4 and the other three states, and it would be tempting to interpret this result as an indication that learners can judge their absolute and certain knowledge in confident relation to other states of knowledge. We might infer, then, that automatic recognition is something of a snap decision for a learner, because it is the one state of knowledge that requires the learner to evaluate their own knowledge in absolute terms. However, we must be cautious and remember that we are dealing with nominal data, so any relationship inferred between the different knowledge states is in need of further exploration and validation. That is beyond the realms of this suite of experiments using a knowledge state rank instrument, simply because there is limited space for exploring every angle of validation.

For the time being, we assume that these results are actually capturing different states of 'threshold meaning knowledge' to paraphrase the way that Schmitt has characterized the VLT (Schmitt, 2000, p.174). Whereas the VLT was thought to

measure such threshold knowledge as its single main construct (but fell somewhat short as the analysis in the previous chapter demonstrated), the knowledge state ranking instrument may be distinguishing between different types of threshold meaning knowledge more finely, if not always in absolutely clear-cut terms. Another possible explanation is that it is easier for learners to distinguish confidently between established knowledge ("I know this verb without thinking") and the three other states of knowledge, but it is much more challenging to make hard and fast distinctions in what they know somewhat, but don't know for certain (i.e., judge the difference between Knowledge States 2 and 3). That said, the need for further validation at a later date is clear.

What we have then is an instrument with (a) a transparent and easy-to-use scoring system, (b) strong reliability, and (c) robust differentiation in how it measures the aspects of automaticity and recognition in relation to the dimension of lexical knowledge. The knowledge state ranking instrument appears to be a very useful tool for testing a large number of items across a diverse population, despite the need for further validation. We have some grounds for believing, then, that we can adapt the knowledge state rank instrument for a further noun recognition test and a verb + noun collocation recognition test.

5.6.2 Learner recognition of lexical verbs

Some interesting but nevertheless elusive patterns started to emerge in which verbs learners recognize most strongly and most weakly. The results from this experiment show that intermediate learners recognize highly frequent lexical verbs most strongly. Only 12 of the 50 most known lexical verbs in Knowledge State 4 are not included in the 2000 most frequent words of English as measured by the BNC (as shown in Table 5.7 below).

Table 5.7 **12 strongly recognized verbs outside the BNC 2000 most frequent range**

Lexical verb	BNC frequency
research	4009
communicate	2933
solve	2262
borrow	2499
project	2872
trust	2130
whisper	2613
obey	4685
debate	3585
stress	2204
feed	2286
function	3927

Conversely, all but 11 of the 50 least known lexical verbs in Knowledge State 4 are outside the 2000 most frequent words of English as measured by the BNC (see Table 5.8 below).

Table 5.8 **11 weakly recognized verbs inside the BNC 2000 most frequent range**

Lexical verb	BNC frequency
emerge	1271
investigate	1804
declare	1583
imply	1731
approve	1780
assume	912
undertake	1695
derive	1836
impose	1517
retain	1547
assess	1570

How might the differences in knowledge of the two sets of verbs be explained? There are different possible explanations, and the question is which explanation is the most plausible here. One difference between the two groups of verbs is that the well-recognised verbs mostly take simple direct object noun phrases, whereas many of the weakly recognised verbs take subordinate clauses rather than simple objects. This takes us towards seeing formal colligation patterns as an influential factor in the

recognition of de-contextualized verbs. However, it seems unlikely that the ease with which learners can generate a formal grammatical relationship for different verbs is going to be decisive.

More probable an explanation is that learners have encountered the most-recognized verbs more often, and, if they have, then they may have encountered them not just as verbs but also as nouns. This second explanation leads us to an important difference between the two sets of verbs: none of the verbs in Table 5.8 can be used as a noun, while, among the most recognized verbs in Table 5.7, many of them can have "multiple class membership" (Biber, Johansson, Leech, Conrad & Finegan, 1999, p.60). They are, in other words, examples of homonymy. Thus, **research**, **project**, **trust**, **whisper**, **debate**, **stress**, and **function** are all verb/noun homonyms. The verb **feed** bears a strong orthographic relationship to the more common noun **food**, and the verb **communicate** should be easily recognizable to learners of English. Since none of the 11 weakly recognized verbs in Table 5.8 is a homonym, the noun for each lexical verb would need to be produced by affixation (i.e., **emergence**, **investigation**, **declaration**, **implication**, **approval**, **assumption**, **undertaking**, **derivation**, **imposition**, **retention**, and **assessment**). We might tentatively infer that many intermediate learners cannot draw on noun knowledge in order to anchor their recognition of such Latinate lexical verbs, whereas the converse would be true with the most-recognised verbs. Since homonymy is a characteristic of frequent vocabulary in English, this explanation seems to fit better as an interpretation. The issue of homonymy is also intriguing with regard to collocation knowledge, so it is one that we will return to in a later experiment when we look at L2 collocation production knowledge.

In the final part of our discussion of the results from the present experiment, we will look more closely at what specific connections can be established between the six source GSL/AWL ranges and learner recognition of lexical verbs.

5.6.3 The relationship between learner recognition and verb frequency

The rank order results show that there are problems in assuming that frequency of occurrence will always strongly correlate with learner recognition of lexical verbs. This can be seen if we look at the distribution of the verbs across the different frequency sources (see Table 5.9 below) and explore the relationship between the

source ranges from the GSL and the AWL, as well as consider British National Corpus ranks for the verbs. The BNC is a useful reference point here because it is based on contemporary frequencies, whereas the GSL was first published in 1936 (Jeffrey, 1952, p.vi), so there must be some doubts as to how well the GSL reflects current usage and frequencies.

When we look at the cumulative distributions in ranges of 10% (from the 10% most recognized verbs through to the 10% least recognized verbs), the first 1000 words of the General Service List prove to be the relatively best predictor of learner recognition—and not unsurprisingly so, as these verbs are highly frequent. What is surprising, however, is that the learner recognition of the verbs from the GSL 1000-2000 range stretches from the higher 0-20% ranges right through to the 0-80% range; this recognition also overlaps with the most common Academic Word List verbs. On the other hand, despite some verbs from the second, third, and fourth Academic Word List ranges in the first 50% of the most-recognised verbs, these three source frequencies tend to cluster towards the 60-100% ranges, or, in other words, the least-recognised verbs.

Table 5.9 Cumulative distributions of source ranges

Note: The first figure in each column indicates the cumulative total; the figure in brackets indicates the increase from the previous percentile range.

	GSL 1-1000	GSL 1000-2000	AWL Range 1	AWL Range 2	AWL Range 3	AWL Range 4
0-10%	9	2	2	1	0	1
0-20%	14 (5)	7 (5)	3 (3)	2 (1)	2 (2)	2 (1)
0-30%	16 (2)	10 (3)	6 (3)	5 (3)	3 (1)	5 (3)
0-40%	19 (3)	12 (2)	11 (5)	6 (1)	6 (3)	6 (1)
0-50%	21 (2)	15 (3)	14 (3)	8 (2)	9 (3)	8 (2)
0-60%	23 (2)	19 (4)	17 (3)	13 (5)	10 (1)	8 (0)
0-70%	24 (1)	21 (2)	20 (3)	17 (4)	12 (2)	11 (3)
0-80%	25 (1)	23 (2)	22 (2)	19 (3)	18 (6)	13 (2)
0-90%	25 (0)	24 (1)	24 (2)	26 (7)	22 (4)	14 (1)
0-100%	25 (0)	24 (0)	27 (3)	26 (0)	27 (3)	20 (6)

At the same time, evidence in favour of using BNC ranks as an alternative selection source is mixed. For example, the 0-2000 BNC range accounts for the 38 of the 50

most-recognised verbs (see Table 5.7 above for the 12 exceptions), whereas the GSL 0-2000 range allows for only 29 of the 50 most-recognized verbs to be accounted for. This is shown in Table 5.10 below. Outside that high frequency core of the BNC, the relationship between learner recognition and frequency would seem to be as varied as with the source ranges used for this study.

Table 5.10 **BNC 2000 and GSL 2000 coverage of the 50 most known verbs (Note: *** indicates the BNC rank is >2000)**

Verb	GSL range	BNC rank		Verb	GSL range	BNC rank
ask	GSL1	154		refuse	GSL1	979
answer	GSL1	989		concentrate		1355
develop	GSL1	411		consider	GSL1	349
write	GSL1	223		remove		891
forget	GSL1	828		hurt	GSL2	1884
build	GSL1	418		avoid	GSL2	866
discuss	GSL2	690		maintain		856
research		***		project		***
communicate		***		trust	GSL1	***
explain	GSL1	537		establish		596
create		482		mention	GSL1	921
break	GSL1	532		comment		1837
select		1632		whisper	GSL2	***
collect	GSL2	1298		obey	GSL2	***
produce	GSL2	332		debate		***
approach		1310		achieve		611
protect	GSL1	1174		identify		797
solve	GSL2	***		occur		650
employ	GSL1	1244		conclude		1764
manage	GSL2	688		stress		***
publish		835		require		343
count	GSL1	1923		feed	GSL1	***
borrow	GSL2	***		include	GSL1	272
focus		1848		function		***

In sum, there are overlaps between the different source ranges used for this research study. The clearest connection between learner recognition and frequency is found with the most frequent 1000 words of English as indicated by the GSL; the GSL 1000-2000 range and the AWL1 range also provide some further evidence of this, whereas learner recognition of verbs in the AWL 2, 3 and 4 ranges is weaker, as

is to be expected. These are broad distinctions rather than rigid demarcations. Finally, the 2000 most frequent words in the BNC can cover 76% of the 50 most known words, whereas the GSL can account for 58% only. We may conclude here that we should in future experiments refer to the BNC and/or some other contemporary corpus for selecting items that are outside the first one thousand most frequent words of English.

5.6.4 Drawing the discussion together

Broadly speaking, this study shows that the knowledge state ranking measure works well, despite the need for further validation at a later date. We have evidence that the instrument produces reliable results for all four states of knowledge with a large population, so we were able to answer the first research question with confidence (*How well does the alternative instrument work?*). As for the second research question (*Which lexical verbs do students know?*), the analysis showed that the first 1000 word range of the GSL was the best predictor of the most known 50 verbs. We also found some evidence that verb/noun homonymy might have some effect on learner recognition. When we explored the third research question (*Is lexical verb knowledge related to lexical verb frequency?*), the results were not so clear. Neither the 2000 most frequent BNC words nor the GSL 1-2000 could cover all of the most recognized lexical verbs (and it is doubtful that we will ever found a perfect correlation between frequency and knowledge).

5.7 Conclusion

In the introduction to this chapter, six areas for improvement were given for designing an alternative measure of lexical verb knowledge. We were concerned with developing an instrument that would:

(i) work from a more specific construct of vocabulary knowledge;
(ii) test more items in less time;
(iii) include both high frequency and academic word list verbs;
(iv) be simpler in its format;
(v) present items independently of each other;
(vi) take explicit account of the use of partial lexical knowledge.

It was important to achieve these improvements because we wanted to settle on a robust instrument that could be used across three interrelated experiments—the verb recognition test reported in this chapter, and the noun recognition test and verb + noun collocation test which will be presented in the next chapter.

We first considered whether a vocabulary scale would offer a viable alternative to the VLT format previously used. Different practical and theoretical problems were noted with the vocabulary scale as an instrument for measuring partial lexical knowledge. On a practical level, we noted that a tool such as the VKS has a confusing scoring system and is time-consuming to use. Theoretically, we questioned the appropriacy of using an instrument that involves distinctly different categories constructed along a single continuum. In the light of these reservations, we decided to opt for a knowledge state ranking instrument that operates along a single dimension and keeps the four states of knowledge separate by differentiating them by either recognition or automaticity. We also took the decision to score each knowledge state separately so that the results would not be contaminated. The knowledge state ranking measure was used to test 149 lexical verbs taken from GSL and AWL source ranges.

The results showed that the instrument is reliable and valid, so we were able to answer the three research questions. Considering the further aim of using a well-piloted and reliable instrument for the next two experiments, the verb recognition test reported in this chapter has been very productive. We now have a tool which we can adapt for other experiments, and, by using a similar kind of measure, we will be able to correlate results from the three experiments. However, before we can run the noun recognition and verb + noun collocation experiments, we need to review the lexical verb database that we have established, so that we can decide on a large number of verb + noun collocations to use. Once we have done that, we will then be able to separate the nouns and run the noun recognition experiment first—and then the collocation test. We have carefully prepared the ground for these two experiments. In the next chapter, we will explore what insights they can offer us for understanding the development of L2 collocation knowledge.

Notes

1. The 1965 article by Dale is discursive in nature and presents this vocabulary scale as part of a discussion of what it means to know a word. As a result, I have

interpreted Dale's discussion as accurately as possible. The original passage reads: "Another source of difference is size of vocabularies discovered as known is influenced by the nature of the test. Knowledge of a word can be placed on a continuum, starting with "I never saw the word before." For example, none of you know the word ***bittles***, ***polentular***, or ***fudular***. They do not exist as meaningful words. Second, we may say, "I know there is such a word but I don't know what it means." Such words might be ***hugger-mugger***, ***adnoun***, ***adit***, ***swingle-tree***, ***détente***, ***antidisestablishmentarianism***, ***laser***, ***serendipity***. A third stage is "a vague contextual placing of the word." You know you make the ***welkin*** ring and that ***hustings*** has to do with elections, but what do these words means specifically? You ***bask*** in the sun. Can you bask in the shade? What are ***kith*** and ***kin***? A ***pied*** piper? Was President Andrew Johnson ***impeached***? These words are in what I call a "twilight zone." Fourth, and finally we reach the stage where we have pinned the word down. We know it. We would recognize it again if we saw it, and we are likely to remember it." (Dale, 1965, p.898)

Chapter 6

Exploring Noun and Verb + Noun Recognition

6.0 Overview

In the previous chapter, we reviewed problems with using Nation's Vocabulary Level Test format for measuring knowledge of individual lexical verbs. We also considered the vocabulary scale as a possible alternative measure, but decided that it would be better to focus on different discrete states of lexical knowledge, i.e.,

1. I have never seen this verb.
2. I have seen but don't know this verb.
3. I know this verb but have to think about it.
4. I know this verb without thinking.

We used these states as the research instrument with 149 lexical verbs drawn from both high frequency sources and the Academic Word List. The results yielded a large database of verbs from which we are now able to choose for a verb + noun (VN) collocation recognition experiment.

In this chapter, we will consider how we can draw on the results from the previous experiment to select lexical verbs for creating a large set of VN collocations. We will then look at testing learner recognition of the nouns included in those VN collocations, before discussing the design of, and results from, a test of collocation recognition. By doing this, we can then compare the results from the three tests: (a) the verb recognition test, (b) the noun recognition test, and (c) the VN collocation recognition test. Such a comparison will let us see in what ways knowledge of individual lexical items (here, verbs and nouns) is connected to knowledge of lexical combinations (here, VN

collocations). The results from the VN recognition test will also provide us with data for examining what makes such collocations relatively easy or difficult to recognize.

6.1 Introduction

The results from the verb test in the previous experiment let us measure how well learners recognized a large number of lexical verbs. The knowledge state ranking instrument proved to be robust and efficient, and verb recognition was found to correlate broadly with a measure of general English proficiency. We carried out the large-scale verb experiment for three main reasons. First, the Vocabulary Levels Test format presented in Chapter 4 was not sensitive to learners' partial knowledge of lexical items. In contrast, the knowledge state ranking format in Chapter 5 allowed learners to make decisions about their perceived degree of knowledge for each verb. Secondly, we wanted to cast the net wider than academic list verbs focused on in Chapter 4 and include a high proportion of frequent verbs. In other words, the verb recognition test laid the necessary groundwork for choosing common everyday and academic collocations. The third reason was that only some of the large pool of verbs will enter into VN combinations. Here, transitivity restrictions and other factors will eliminate some verbs from the candidate pool. So, testing a large number of lexical items in the previous experiment now affords us some flexibility in choosing items for the collocation recognition and noun recognition tests presented in this chapter.

The three tests of recognition—Verb, Noun, and VN—form a series of interrelated experiments. In this chapter, we will refer to the verb database from Chapter 5 as we make the selection of verbs for the VN test. We will then create a set of VN collocations and isolate the nouns from the VN recognition test for the noun test, before analyzing the results from that test. Finally, we will compare the

results of the VN recognition to those of the two other tests.

6.2 Research questions

Two separate sets of research questions are addressed in this chapter, as well as one question based on all three recognition experiments:

Noun recognition test

1) How well do learners recognize nouns?
2) Which nouns are better known than others?
3) Does noun knowledge correlate with general English proficiency?

VN recognition test

4) How well do learners recognize collocations?
5) Which collocations are better known than others?
6) Does collocation knowledge correlate with general English proficiency?

All three recognition tests

7) Do V knowledge, N knowledge, and VN collocation knowledge correlate with each other?

6.3 Research design and method

The selection of verbs for the VN test raises two general questions: Which verbs should be chosen, and how many? Several constraints impinge on the answers, such as length and time of the test, the inclusion of mis-collocations as a validity check, and the availability of noun collocates for different verbs. We will consider each of these constraints in more detail.

With regard to the length and time of the test, the intended purpose of the VN experiment was to test a large number of items—to include about 100 or so collocations, with more than one noun collocation for each verb. Because the VN collocation test involved two-word items, it did not seem reasonable to aim for another 150-item test:

The students would need more time to read the two-word combinations. I judged that a 120-item test could be completed in about 10 minutes.

Having decided the total number of items, I was faced with the second constraint of making some of the items mis-collocations. These mis-collocations might provide a useful way to check the validity of the knowledge-state rating instrument. With a 120-item collocation test, I wanted to use 40 verbs from the previous verb test, which meant that each verb would feature three VN collocations. However, the inclusion of mis-collocations as a validity check would reduce the number of real collocations. At first, I considered making one collocation in each set of three VN collocations a mis-collocation, but I felt that this risked making the test transparent and reducing it to one of guessing which of any three collocations was false. For this reason, I limited the number of mis-collocations to 20, leaving a total of 100 real collocations. In other words, the test consisted of 20 sets of three real collocations and 20 sets of two real collocations + one mis-collocation.

As for which 40 verbs should be chosen, I decided this by looking more closely at the results from the verb test. I felt that it was important to choose strongly recognized verbs, as well as some more weakly recognized verbs. I also needed to check any selected verbs for their possible collocates. To make the selection, I took the combined totals for Knowledge States 3 and 4 for the verbs (see Chapter 5, Table 5.6). I also considered their BNC rank, because the results from the previous experiment suggested this rank is a better indicator of frequency than the General Service List outside of the first 1000 words of the GSL. To cross-check the verbs for their noun collocates, I consulted the Cobuild Bank of English (HarperCollins, 1995a), which led to further constraints. Among the restrictions in the choice of verbs, lack of transitivity was the first to exclude verbs such as **immigrate** and **occur**. A second restriction

resulted from some verbs having predominantly adjectival forms as their most frequent collocation pattern. This was the case with **perceive/perceived**, for example. Another type of restriction came up with **ask**, which did not have sufficient different noun collocates:

ASK 32568
questions 2068, question 1262, people 1023, whether 778
ASKED 60302
whether 1792, questions 1346, question 1238
ASKING 13220
questions 998, people 605
ASKS 4657
whether 199, question 148, questions 148

Ask had just two noun collocates available, both of which were high frequency, and it seemed doubtful whether such items would lead to any variation in learner recognition. It was therefore dropped.

However, even when verbs such as **employ** had a clear set of noun collocates, many of these noun collocates seemed remarkably similar. This proved to be another restriction. **Employ** occurred in its base form a total of 2353 times with 261 occurrences for **people**, 68 for **workers**, 60 for **staff**, 33 for **women**, whereas it had 29 occurrences for **means**, 29 for **techniques** and 21 for **methods**. These collocation patterns were also evident in its other forms:

EMPLOYED (5727 total occurrences)
people 425
EMPLOYING (1259 total occurrences)
people 188, staff 45, workers 42, women 28, techniques 20
EMPLOYS (1252 total occurrences)
people 359, workers 63, staff 54

Although such semantic overlap in collocation ranges is a fundamental feature of collocation, it made it difficult to consider a verb like **employ** as a viable item.

Despite starting from a large pool of candidate verbs from the verb test, the design of the VN collocation test was affected by the constraints discussed above. One result was that some nouns occurred more than once in the final selection (for example, **create problem** and **explain problem**). The final selection of 40 verbs is shown in Table 6.1.

Table 6.1 **40 verbs selected for the VN collocation test**

Verb	Total KS3+4	BNC rank	Verb	Total KS3+4	BNC rank
change	217	154	approve	172	1780
protect	216	1174	assume	169	912
solve	215	2262	construct	162	2225
break	214	532	preserve	162	2363
create	214	482	declare	158	1583
explain	213	537	consume	156	4841
conclude	208	1764	investigate	155	1804
defend	206	2182	regulate	152	4055
adopt	205	1186	predict	151	2396
recommend	204	1618	evaluate	151	3254
observe	204	1323	dismiss	144	2148
seek	202	624	distribute	142	2787
promote	202	1571	seize	135	3002
occupy	197	2021	restrict	117	2193
acquire	193	1512	impose	109	1517
threaten	189	1475	derive	103	1836
define	188	1097	retain	91	1547
conduct	187	1835	compensate	85	4603
obtain	182	824	assess	77	1570
govern	175	2938	implement	43	2243

Of the 40 verbs, 24 came from the BNC top 2000. A further 11 came from the BNC 2001-3000 range, and five came from the 3-5K BNC range.

6.3.1 Design of the VN collocation test

Once I had chosen three noun collocates for the 40 verbs, the next step was to create 20 mis-collocations, each one occurring once with 20 of the 40 different verbs. The false collocations needed to be unevenly spread across the test, so that there was no regular pattern to their location. I made these decisions largely intuitively as to how the mis-collocations should be created, although, where possible, a noun collocate from another verb was used. The final set of mis-collocations is shown in Table 6.2 below.

Table 6.2 20 mis-collocations for the VN collocation test

adopt profit	derive conversation	preserve response
approve opportunity	dismiss willingness	promote sense
change memory	explain address	regulate scene
conclude price	impose movement	seek holiday
construct benefits	investigate opinion	seize progress
create temperature	occupy deal	solve peace
defend series	predict territory	

In creating the mis-collocations, I was working from the assumption that recognition of a particular VN combination depends on how often a combination has been met. Learners would not have met any of the false collocations before and so could be expected to rate them weakly, whereas they might have previously met (or not) real collocations and so should vary more in how strongly they recognized them.

The full set of 120 VN collocations is shown in Table 6.3 below, with mis-collocations marked in bold italics. The collocations are arranged alphabetically by verb.

Table 6.3 **120 VN collocations selected (with mis-collocations shown in bold italics)**

acquire	skills		***explain***	***address***
acquire	reputation		explain	problem
acquire	weapons		explain	reasons
adopt	approach		govern	country
adopt	child		govern	relations
adopt	***profit***		govern	universe
approve	budget		implement	changes
approve	***opportunity***		implement	plan
approve	plan		implement	policies
assess	damage		impose	ban
assess	risk		***impose***	***movement***
assess	situation		impose	sanctions
assume	importance		investigate	allegations
assume	responsibility		investigate	cases
assume	role		***investigate***	***opinion***
break	ground		observe	behaviour
break	record		observe	ceasefire
break	rules		observe	silence
change	***memory***		obtain	approval
change	mind		obtain	copy
change	policy		obtain	permission
compensate	lack		occupy	mind
compensate	loss		***occupy***	***deal***
compensate	victims		occupy	space
conclude	agreement		predict	future
conclude	***price***		***predict***	***territory***
conclude	talks		predict	victory
conduct	investigation		preserve	life
conduct	orchestra		***preserve***	***response***
conduct	research		preserve	status
construct	***benefits***		promote	peace

construct	model		promote	programme
construct	theory		***promote***	***sense***
consume	alcohol		protect	body
consume	energy		protect	environment
consume	quantities		protect	interests
create	jobs		recommend	action
create	problem		recommend	book
create	***temperature***		recommend	changes
declare	independence		regulate	levels
declare	support		***regulate***	***scene***
declare	war		regulate	use
defend	decision		restrict	ability
defend	***series***		restrict	access
defend	title		restrict	use
define	problem		retain	control
define	role		retain	moisture
define	word		retain	place
derive	benefit		seek	advice
derive	***conversation***		seek	election
derive	pleasure		***seek***	***holiday***
dismiss	claims		seize	opportunity
dismiss	reports		seize	power
dismiss	***willingness***		***seize***	***progress***
distribute	food		solve	crisis
distribute	money		solve	mystery
distribute	leaflets		***solve***	***peace***
evaluate	work		threaten	future
evaluate	performance		threaten	stability
evaluate	situation		threaten	strike

6.3.2 Design of the noun recognition test

Once I had chosen the VN collocations, I could construct the noun recognition test.

Here, I adjusted the same four states of knowledge from the previous verb test:

1. I have never seen this noun.
2. I have seen but don't know this noun.
3. I know this noun but have to think about it.
4. I know this noun without thinking.

6.3.3 Research method

The 120-item noun test was administered to four class groups or a total of 93 students. These groups were Area Studies (n=10), Environmental Science (n=28), Humanities (n=29) and Medical Science (n=26). Three of these groups were undergraduates and had been reliably distinguished by an in-house English proficiency placement test. Medical Science was upper-intermediate, Humanities intermediate, and Environmental Science low-intermediate. Area Studies, the postgraduate class, ranged from advanced to intermediate. All 93 students took the three recognition tests (V, N, and VN). We will next look at the results for the noun test, after which the results for the VN collocation test will be reported.

6.4 Results for the noun recognition test

The data for the noun recognition test were analysed in two ways. For word-based analysis, a cumulative scoring of learner recognition was conducted. Here, all the learner claims for Knowledge State 4 ("I know this noun without thinking") were first scored as 1, with all other Knowledge States scored 0. Next, all the learner claims for Knowledge States 3 and 4 ("I know this noun without thinking" & "I know this noun but have to think about it") were scored as 1, with the other two Knowledge States

scored 0. These two sets of scores were then used to produce rank order listings of the nouns for both Knowledge State 4 and for Knowledge States 3 and 4 combined. Secondly, in terms of subject-based analysis, an analysis of variance was carried out with the two sets of recoded scores to see if there were significant differences between the 4 class groups. These analyses allow us to see clearly how confidently learners claim to recognise nouns, as well as to establish if there any correlations between general English proficiency levels and noun recognition.

6.4.1 Noun recognition: descriptive statistics

Noun recognition proved to be very high as the descriptive statistics in Table 6.4 below indicate. In Knowledge State 4, the mean total score of 110.41 (out of a total of 120), together with the very low s.d. of 5.29, underline just how strongly nouns were claimed to be known. The corresponding results for Knowledge States 3 and 4 combined (total 113.59, s.d. 3.41) further point to a possible ceiling effect in noun recognition.

Table 6.4 Nouns assigned to different knowledge states

	Knowledge State 4	Knowledge States 3 & 4
N	93	93
Mean	110.41	113.59
Standard deviation	5.29	3.41
Cronbach's α	.86	.79

The internal consistency of the noun recognition is high for Knowledge State 4 at .86, but declines to .79 when Knowledge State 3 is combined, which we may attribute to the increasing lack of variance in the second set of recoded scores. The small s.d. values also point to the lack of variance.

6.4.2 Noun recognition: word-based analysis results

Overall, only 13 nouns were recognized by fewer than 75 students or Knowledge State 4, with just seven of these nouns (**ban, stability, reputation, leaflets, ceasefire, sanctions,** and **allegations**) remaining unrecognized in Knowledge States 3 & 4 combined. These results are shown in Table 6.5 below.

Table 6.5 Number of subjects rating nouns in Knowledge State 4 and Knowledge States 3 & 4 combined, with total <74 (N=93)

Noun	Total KS 4	Total KS 3 & 4
moisture	72	82
programme	72	78
investigation	68	79
willingness	66	86
approval	64	82
budget	64	79
ban	59	70
stability	58	71
reputation	38	56
leaflets	25	30
ceasefire	10	20
sanctions	8	20
allegations	3	15

In brief, the recognition of nouns was so high that there was little overt variation in the scores except with a small number of nouns.

6.4.3 Noun recognition: subject-based analysis results

A one-way ANOVA to compare means between the groups for Knowledge State 4 was performed and showed significance ($F(8,89) = 6.401$, $p<.001$). A post-hoc Scheffe

test showed a significant difference between the upper-intermediate Medical Science group and the low-intermediate Environmental Science group ($p = .001$). For Knowledge States 3 & 4 combined, a one-way ANOVA also showed significance ($F(3,89) = 3.503$, $p<.05$). A post-hoc Scheffe test showed a significant difference between the same groups as for Knowledge State 4, Medical Science and Environmental Science, ($p = .030$).

6.4.4 Drawing together the noun recognition results

The noun recognition test produces three main findings. The first is that the students claim very high recognition of nouns. The second is that noun recognition shows limited significant correlation with broad differences in general English proficiency level (see 6.4.3). Just as the verb recognition did, the noun test distinguishes between upper intermediate and lower intermediate bands only. Despite the significant difference between two groups, the third important result is that the noun recognition test indicates that very few of the nouns were less than highly recognized. There seems, then, to be a ceiling effect, which may affect our later analysis of the interaction between verb knowledge and noun knowledge in relation to collocation knowledge. We will now look at the method and results for the VN recognition test.

6.5 Method for the VN recognition test

The knowledge states for the VN collocation test were slightly adapted from the previous noun and verb tests. The four states were:

1. I don't know this combination at all.
2. I think this is not a frequent combination.
3. I think this is a frequent combination.
4. This is definitely a frequent combination.

6.6 Results for the VN recognition test

The students were given no indication that some of the 119 collocations were false. (One real collocation was unfortunately lost in page formatting.) The test was administered to the same four class groups (N=93).

The data for the 119 VN collocation recognition test were scrutinized in the following ways. The real collocations were separated from the mis-collocations, so that each set of items would not interfere with each other. As with the noun test, the scores for the real collocations were progressively recoded by starting from Knowledge State 4 and by then adding in the claims of recognition from Knowledge State 3. These two sets of scores were then used to produce rank order listings of the real collocations for (a) Knowledge State 4 and (b) Knowledge States 3 and 4 combined.

For a subject-based analysis, three operations were carried out. First, a series of 2 x 4 chi-square analyses was completed on all 99 collocations for the recoded Knowledge States 4 data. This was done in order to establish whether there were significant associations between collocation knowledge and verb and noun knowledge. The second subject-based analysis involved comparing each student's totals for known verbs and known nouns with known collocations, in order to establish possible correlations. Finally, a comparison of groups was done to test for correlations between collocation knowledge and general English proficiency.

In contrast, the mis-collocations were separately recoded for each knowledge state. A series of tests was then run on these four sets of data to investigate whether the mis-collocations maintained a similar degree of recognition or not in each state.

In sub-sections 6.6.1 – 6.6.6, we will look at the results for the real collocations, before we turn to the findings for the mis-collocations in section 6.7.

6.6.1 Real collocations: descriptive statistics

The descriptive statistics for the real collocations are shown below in Table 6.6 for Knowledge State 4 and for Knowledge States 3 and 4 combined.

Table 6.6 Descriptive statistics for real collocation recognition

	KS 4	KS 3 & 4
N	93	93
Minimum	0	4
Maximum	80	90
Mean	30.49	53.71
Standard deviation	19.14	17.68
Cronbach's α	.96	.95

For Knowledge State 4, the mean score is 30.49 (out of a total of 99), which increases substantially for combined Knowledge States 3 & 4 (53.71). The internal reliability of each set of scores, as measured by Cronbach's alpha, remains almost consistent, as do the relatively high s.d. values.

6.6.2 Real collocations: rank order results

The rank order results for the real collocations are shown in Table 6.7 on the next two pages, where the collocations have been ordered from most to least known for Knowledge State 4.

Table 6.7 **Rank order results for real collocations for Knowledge State 4 and Knowledge States 3 & 4 combined (N=93)**

Collocation	KS 4	KS 3&4	Collocation	KS 4	KS 3&4
change mind	79	92	obtain approval	24	51
protect environment	74	85	regulate use	24	44
explain reasons	72	85	derive benefit	23	42
predict future	67	83	construct model	22	57
protect body	67	86	define role	22	57
recommend book	67	82	evaluate situation	22	44
govern country	65	84	obtain permission	22	54
declare independence	62	75	recommend changes	22	55
break rules	61	78	restrict ability	22	45
explain problem	61	79	approve plan	21	49
break record	60	79	govern universe	20	48
threaten future	56	78	preserve status	20	55
conduct orchestra	53	71	assume responsibility	19	46
occupy space	53	76	defend decision	19	41
solve mystery	52	76	protect interests	19	39
observe behaviour	51	78	assess risk	18	40
change policy	50	78	assume importance	17	41
define word	50	71	compensate loss	17	36
consume energy	46	66	approve budget	15	32
defend title	44	70	obtain copy	15	34
adopt child	43	59	retain moisture	15	41
declare war	43	68	assess damage	14	30
occupy mind	42	71	compensate lack	14	31
promote peace	41	69	dismiss claims	14	39
evaluate work	38	64	observe silence	14	33
evaluate performance	38	63	retain control	14	44
preserve life	38	58	assume role	13	30
construct theory	36	69	consume quantities	13	32
investigate cases	35	56	derive pleasure	13	35
predict victory	35	69	dismiss reports	13	38
seek advice	35	65	impose ban	13	25

solve crisis	33	54	seek election	13	36
conclude agreement	32	64	acquire weapons	11	27
consume alcohol	32	54	assess situation	11	38
seize power	32	56	compensate victims	10	31
distribute food	31	48	govern relations	9	36
promote programme	30	59	implement plan	9	21
acquire skills	29	59	acquire reputation	8	23
seize opportunity	29	50	declare support	8	42
conduct research	28	53	distribute leaflets	7	25
restrict access	28	47	implement changes	7	18
restrict use	28	49	observe ceasefire	7	15
threaten stability	28	48	adopt approach	6	24
create jobs	27	50	implement policies	6	23
recommend action	26	44	break ground	5	14
define problem	25	58	investigate allegations	5	20
distribute money	25	51	threaten strike	5	19
regulate levels	25	48	retain place	4	27
conduct investigation	24	46	impose sanctions	3	15
create problem	24	41	--	-	-

A number of very general points can be made about what this first level of analysis reveals. Very few collocations are recognized with certainty. If we follow the KS4 column down as far as **defend title**, we find just 25 collocations are definitely claimed to be frequent by more than half the students. Of these 25, there are four pairs of collocations which involve the same verb: **change mind/change policy, protect environment/ protect body, explain reasons/explain problem,** and **break rules/break record.** These verbs are among the six most recognized lexical verbs, as shown in Table 6.1. The other 17 collocations in the top 25 involve a different verb each time. On the other hand, if we look at the 30 least recognized collocations, we may observe that 12 involve six verbs that are among the 10 least recognized verbs

(again, see Table 6.1): **compensate lack/victims, derive pleasure, dismiss claims/reports, retain moisture/control/place, implement plan/changes /policies**, and **impose ban/sanctions**. To gain a finer understanding of the collocation results, we need to look more closely at the strength of association between knowledge of the two individual components, the verb and the noun, with the collocation.

6.6.3 Real collocations: chi-square analysis results

To see whether there was a significant association between collocation recognition and verb and noun knowledge, a series of 2 x 4 chi-square analyses for collocation recognition and verb and noun knowledge was completed on the 99 real collocations for the recoded scores in Knowledge State 4.

An example analysis is shown in Table 6.8 below. The row "Collocation (1)" gives the results for **acquire reputation** claimed as recognised. The row "Collocation (0)" shows the results for the collocation claimed as unrecognised. The columns categorise verb and noun knowledge: "VN" shows the results for learners claiming that they knew both the verb and noun, "Vn" just the verb, "vN" just the noun, and "vn" neither the verb nor the noun.

Table 6.8 **Example chi-square results: acquire reputation**

	VN	Vn	vN	vn	Pearson 93 cases
Collocation (1)	5	2	1	0	X^2=12.33, d.f.=3,
Collocation (0)	29	34	3	19	p= .0151

Table 6.8 indicates that five students knew both verb and noun *and* recognised the collocation, but 29 students didn't recognise the collocation, despite knowing both the verb and noun. Two students who recognised the collocation knew the verb but not the noun; one student who recognised the collocation knew the noun but not the verb, and so on. The association between the four categories of verb/noun knowledge and collocation recognised/not recognised is, in this case, significant ($p = .0151$). (See Appendix 6.1, for each separate chi-square analysis.)

In total, 35 collocations showed significance of association ($p < .05$). These are reported in Table 6.9.

Table 6.9 Collocations showing significance of association

acquire reputation, approve budget, approve plan, assess damage, compensate lack, compensate loss, compensate victims, conclude agreement, conduct orchestra, declare independence, declare war, define role, define word, dismiss claim, distribute food, distribute leaflets, govern country, implement plan, implement policies, impose ban, investigate cases, observe behaviour, obtain copy, obtain permission, occupy mind, occupy space, preserve life, protect environment, recommend book, restrict access, restrict use, retain place, seek advice, seize opportunity, seize power

6.6.4 Real collocations: collocation recognition and verb/noun knowledge results

To explore more closely the relationship between collocation recognition and verb/noun knowledge, the raw frequencies for VN knowledge were looked at. Logically, we would not expect a collocation to be recognized if either the verb or noun is not known.

However, in a small number of cases, vN, Vn, and vn claims were greater than VN knowledge. These results are shown in Table 6.10.

Table 6.10 Vn, Nv, and vn is greater than VN for recognised collocations

Vn > VN (3 cases)	vN > VN (11 cases)	vn > VN (3 cases)
evaluate work (24 : 9)	assess risk (12 : 6)	investigate allegations (2 : 1)
investigate allegations (2 : 1)	assess situation (6 : 5)	investigate cases (9 : 1)
observe silence (10 : 3)	derive benefit (13 : 10)	retain place (3 : 0)
	distribute money (13 : 12)	
	implement changes (6 : 1)	
	implement plan (6 : 3)	
	implement policies (4 : 2)	
	recommend action (54 : 22)	
	retain control (10 : 4)	
	retain moisture (10 : 5)	
	retain place (1 : 0)	

In most of the 17 "aberrant" cases shown in the table above, the differences are very small. The two exceptions are **evaluate work** and **investigate cases**. In 81 cases, VN knowledge was greater than Vn, Nv, and vn claims. In one case, VN was equal to vN. These three results (81, 17, and 1) cover the 99 real collocations. However, there were only 22 collocations where students claimed knowledge of both verb and noun, as well as the collocation, as shown on the next page in Table 6.11. It is the top 15 of this group of collocations that we can characterize as better known than the others in that more than half of the students recognize them.

Table 6.11 **Number of subjects knowing verb, noun and collocation**

Collocation	Known	Unknown
change mind	79	14
protect environment	73	17
explain reasons	70	19
protect body	66	23
explain problem	60	28
break record	59	31
govern country	59	17
break rules	58	32
change policy	50	43
threaten future	50	31
declare independence	49	14
observe behaviour	49	32
predict future	49	16
occupy space	48	27
solve mystery	47	37
define word	43	28
occupy mind	38	32
declare war	35	28
consume energy	34	19
conduct orchestra	31	17
preserve life	29	22
distribute food	15	11

In general, the results from the collocation test indicate that knowledge of individual verbs and nouns does not necessarily entail recognition of their combination in a VN collocation.

6.6.5 Real collocations: comparison of groups results

For Knowledge State 4, a one-way ANOVA failed to show significance between any groups. No correlation was established between general English proficiency and collocation recognition in Knowledge State 4.

6.6.6 Drawing together the real collocation recognition results

Three interesting points emerge from the real collocation recognition results. The first is that the test instrument shows high reliability and also produces greater variation than the noun test. Here, we can note that the mean scores for collocation recognition are relatively low for Knowledge State 4 and Knowledge States 3 and 4 combined. The second point is that we could find significance of association in just over a third of the 99 collocations. As shown in Table 6.9, many of these collocations involve academic word list verbs and nouns where student knowledge is limited. However, this group of collocations also includes more 'everyday' collocations such as **seek advice** and **seize opportunity**, which were well-recognised. These two contradictory tendencies in the test of association make it difficult to draw meaningful insights, unless we interpret the significance of association as characterizing the relationship as either "very likely" or "very unlikely." The final point is that collocation recognition lags behind knowledge of individual verbs and nouns. Here, we were able to identify just 15 out of 99 collocations where learner recognition was higher than verb and noun knowledge across more than half of the subjects. We will consider this issue in the discussion section, but we will briefly examine, first, the results with the 20 mis-collocations, and, second, the correlations between the three recognition tests.

6.7 Results: mis-collocations

Included in the collocation recognition test were 20 mis-collocations intended to act as a reliability check. Logically, a mis-collocation should be mainly scored in Knowledge State 1 ("I don't know this combination at all.") or 2 ("I think this is not a frequent combination.") because learners would have never met it. If this were so, then we would also expect low reliability co-efficients, as scores would be mainly clustering in those two states. The descriptive statistics for the mis-collocations are presented below in Table 6.12.

Table 6.12 Mis-collocations assigned to each knowledge state

	KS1	KS2	KS3	KS4
N	93	93	93	93
Mean	5.70	7.53	4.46	2.20
Standard deviation	5.83	4.44	3.89	3.11
Cronbach's α	.93	.80	.81	.84

The descriptive statistics show limited variation in internal consistency, and the alpha co-efficient does not drop away from Knowledge States 2 to 4. The s.d. values are also relatively high given the small number of items involved.

Table 6.13 below reports how individual mis-collocations were scored in each knowledge state. The mis-collocations are arranged in descending order for the total of Knowledge States 1 and 2 combined. This score represents how many students rejected each mis-collocation. The dotted line across the table separates those mis-collocations rejected by more than two thirds of the students.

Table 6.13 Mis-collocations claimed as known

Mis-collocation	KS4	KS3	KS3+KS4	KS2	KS1	KS2+KS1
create temperature	7	11	18	40	35	75
solve peace	7	17	24	36	33	69
regulate scene	6	17	23	44	25	69
occupy deal	6	18	24	43	25	68
defend series	7	18	25	37	31	68
dismiss willingness	13	15	28	28	37	65
seek holiday	11	14	25	39	26	65
derive conversation	7	21	28	30	35	65
promote sense	6	23	29	33	31	64
adopt profit	6	24	30	32	30	62
predict territory	10	22	32	35	26	61
seize progress	7	24	31	28	31	59
impose movement	12	22	34	26	33	59
conclude price	11	23	34	38	20	58
approve opportunity	7	28	35	38	20	58
preserve response	9	26	35	40	18	58
change memory	16	23	39	36	18	54
construct benefits	13	26	39	35	19	53
explain address	27	16	43	31	19	50
investigate opinion	17	27	44	31	18	49

The results show that, for all 20 mis-collocations, more than half the learners rejected the combinations as likely to occur. Only 10 of the 20 mis-collocations were clearly rejected by more than two thirds of the students. Given also the weak rate of real collocation recognition earlier reported, we can conclude that the mis-collocations used in this test are neither uniformly rejected nor particularly useful as a validity check.

Having reported the results for the mis-collocations, we will next have a look at possible correlations between the three recognition tests.

6.8 Real collocations: correlations between verb, noun, and collocation totals for Knowledge State 4

A comparison of individual scores for verb, noun, and collocation knowledge was carried out. Kendall's tau-b correlation was performed on the total scores for the three recognition tests in Knowledge State 4 and showed significance:

- collocation-noun $r = .324$; $N = 93$; $p < 0.001$
- collocation-verb $r = .400$; $N = 93$; $p < 0.001$
- verb-noun $r = .749$; $N = 93$; $p < 0.001$.

The verb test and the noun test show the strongest correlation and produce a scatter that is moderately distributed from the diagonal (see Figure 6.1 below).

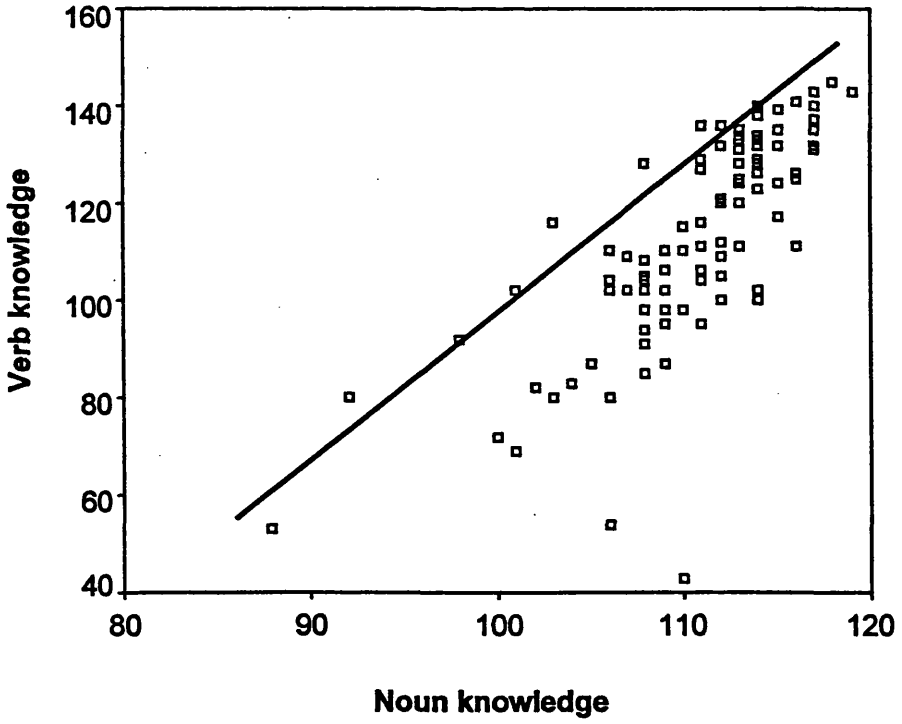

Figure 6.1 **Scatterplot for the correlation between noun and verb knowledge**

Although the correlations between the collocation test and the individual verb and noun tests are weaker, they are nevertheless significant.

6.9 Discussion

We will start the discussion by considering how well we have been able to answer the research questions presented in Section 6.2. We will then narrow down our focus to two specific issues:

- What might underlie strong collocation recognition?
- How can we characterize collocation knowledge, given that we have not been able to establish a significant relationship with overall English proficiency?

The results from the noun and collocation recognition tests enable us briefly to address the research questions presented in Section 6.3 of this chapter. These are:

1) How well do learners recognize nouns?
2) Which nouns are better known than others?
3) Does noun knowledge correlate with general English proficiency?
4) How well do learners recognize collocations?
5) Which collocations are better known than others?
6) Does collocation knowledge correlate with general English proficiency?
7) Do verb knowledge, noun knowledge, and verb + noun collocation knowledge correlate with each other?

We have established that learners have a very high recognition rate for nouns (*Research Question 1*), with just 13 nouns less than fully recognized (*Research Question 2*). We have also shown that noun recognition, like verb recognition, significantly correlates with broad bands of general English proficiency (*Research Question 3*). In contrast, we found that collocation recognition does not show significant correlations with English proficiency (*Research Question 6*). In observing that collocation recognition lags behind recognition of the individual nouns and verbs forming a given collocation

(*Research Question 4*), we were able to identify a set of 15 collocations that were strongly recognized (*Research Question 5*). Finally, we were able to establish that the three recognition tests show significant correlations (*Research Question 7*). That is, knowledge of individual verbs and nouns has an important but weak relationship to knowledge of verb + noun collocations.

How we might interpret that relationship is another matter, however. One way is to take the 15 most recognized collocations (see Table 6.11) and to profile the lexical frequency of both elements. This will let us see whether learners are strongly recognizing collocations where both the verb and the noun are frequent. "Frequent" here means within the first 2000 words of English. A lexical profile of the 30 words used in those 15 collocations (Cobb, 2003) shows that just four words are outside the high frequency range: **environment, occupy, policy,** and **predict** (see Table 6.14 below).

Table 6.14 Lexical profile of the 15 most recognized collocations

Range	Item
K1	body, break, change, country, declare, explain, future, independence, mind, observe, problem, protect, reasons, record, rules, space
K2	behaviour, govern, mystery, solve, threaten
AWL	environment, occupy, policy, predict

The lexical profile indicates that the most recognized collocations predominantly consist of highly frequent words.

A second way to interpret the relationship is to ask whether the most recognized collocations are lexically congruent with their equivalent translations in Japanese. The

15 collocations were translated into Japanese and then presented to three Japanese specialists of English language and literature (K, M, and S) to rate independently for the lexical congruence between English and Japanese. Where they judged lexical congruence to hold, they classified the combinations as "A." If they judged either the verb or noun or verb to be lexically incongruent, they were asked to categorise the collocation as "B." Where they judged neither "A" nor "B" to hold, the interraters were invited to code the item as "C." The results are shown in Table 6.15 below. Only one collocation was judged by one rater to be lexically incongruent (**occupy space**), and one of the raters (K) evaluated **change mind** as having a different translation in Japanese from the one given on the rating sheet.

Table 6.15 Lexical congruence of the 15 most recognized collocations

Collocation	K	M	S
change mind	C	A	A
protect environment	A	A	A
explain reasons	A	A	A
protect body	A	A	A
explain problem	A	A	A
break record	A	A	A
govern country	A	A	A
break rules	A	A	A
change policy	A	A	A
threaten future	A	A	A
declare independence	A	A	A
observe behaviour	A	A	A
predict future	A	A	A
occupy space	B	A	A
solve mystery	A	A	A

In short, the most recognized collocations show a high frequency lexical profile and almost exact lexical congruence between English and Japanese.

Our final area of discussion concerns how we might best characterize collocation knowledge, given that we have not been able to establish a significant relationship with overall English proficiency. As we saw in Chapter 2, most of the previous studies into L2 collocation knowledge have tended to work with advanced learners (see Table 2.4). Only Bonk (2000), Gitsaki (1996), Chi, Wong and Wong (1994), and Mochizuki (2002) used exclusively non-advanced learners. All four studies found that low-level learners have collocation knowledge, even if their level of knowledge is limited. Through the three interrelated experiments into V, N, and VN recognition, we have come a little further than those previous studies. We have established that intermediate learners' collocation knowledge is not guaranteed even if they have strong recognition of the individual components of particular collocations. This suggests that an intermediate Japanese learner, in as far as our sample is representative, does not typically learn lexical combinations first and then analyse them later. All the evidence that we have from these experiments points to collocation development proceeding analytically by acquisition of individual lexical items. We have also confirmed that some intermediate learners do recognise collocations strongly, so we cannot characterize collocation as "an advanced type of vocabulary knowledge" as Schmitt does (2000, p.89). In fact, because the results show that strongly recognised collocations have a highly frequent lexical profile, it is quite likely that intermediate learners can both recognize and produce—a limited range of—collocations in English. It may be simply the case that we have not yet found a way to observe this properly, so collocation as advanced L2 knowledge may be viewed as a research default position,

faute de mieux. This is one of the key questions that we will pursue further in the next chapter when we take a closer look at collocation production.

6.10 Conclusion

In this chapter, we completed a suite of three lexical recognition experiments, using a knowledge-state rating instrument. In each case, the instrument was adapted and proved to be reliable and consistent, so we may have some confidence in the results obtained. We were careful to measure learner recognition of individual lexical verbs and nouns before measuring recognition of collocations composed of those twin elements. This allowed us to correlate the three measures and to ascertain that, although collocation knowledge is related to knowledge of the lexical items that form a VN combination, it is also weaker and lags behind knowledge of individual lexical items.

The further analysis of the 15 most recognized collocations suggested that lexical frequency and congruence may play a role in collocation recognition. However, one limitation of this experiment is that we could draw on only a very small number of items in making such a suggestion. A second limitation was that, although we found that collocation recognition did not correlate significantly with overall English proficiency, we could not explain why this is so.

This leaves us with a number of interesting paradoxes. First, we have shown that intermediate learners have a restricted level of second language collocation knowledge, but we have been unable to connect this to overall English proficiency. Next, we have suggested that the L2 mental lexicon works in part from individual lexical items rather than lexical combinations, but we do not have any means at present of understanding this further. Finally, we have shown that intermediate learners can

recognize collocations, but we do not know empirically whether they can also produce them.

It is this last question that we will turn our attention to in the next chapter. We have completed four experiments on recognition in Chapters 4, 5, and 6; it is now time to switch our focus and investigate how collocation production can be measured. This is what we will explore in the next chapter.

Chapter 7

Investigating Individual Variation in Collocation Production over Time

7.0 Overview

In this chapter, we shift our focus from investigating L2 collocation recognition to exploring L2 collocation production. I begin by reviewing a number of problems with previous investigations of productive L2 collocation knowledge and by considering how we can adapt a measure from word association studies to investigate the development of individual collocation knowledge over one academic year. I present the design of the adapted measure and then explain how it was administered six times over one academic year, with the learners required each time to write down three word associations and collocations in response to a small number of stimulus nouns. I then report the results in detail and discuss individual variation for both word associations and collocates. Although this measure of L2 collocation knowledge suffers from some weaknesses, I conclude that an instrument of this kind, once further developed, may help us understand L2 collocation production better than conventional tests narrowly focussed on identifying erroneous collocation use.

7.1 Introduction

As we saw in Chapter 2, previous investigations of productive L2 collocation knowledge have involved single interventions focused on identifying erroneous L2 collocation use at the sentence or text level through translation, cloze and/or error analysis. The limited scope of such earlier studies is problematic for a number of reasons. First, confining the elicitation of collocates to one correct answer does not show what other collocates learners may know for a particular target word. Second,

although corpus-based measures allow for more instances of a particular type of collocation to be recovered, they tend to generalize across individual productions towards standard types of error for a particular group of learners. This obscures insights into individual variation in collocation production and tells us a great deal more about what groups of learners do not know rather than what individuals do. Third, an experimental measure such as Schmitt's limits the amount of data that can be yielded because the learners are required to produce sentences rather than individual collocates to show their collocation knowledge. Finally, nearly all of the previous research reviewed has featured single interventions, so we have no sense of the specific development of individuals' L2 collocation knowledge over time. There are then questions as to how we can best observe L2 learners' collocation knowledge in a time-efficient and effective manner. It may, however, be possible to adapt from L2 word association studies a measure that meets some of the concerns that we have just outlined and that also allows us to examine L2 collocation knowledge longitudinally.

Word association studies are potentially useful in investigating L2 collocation knowledge for two basic reasons. First, experimental L2 word association studies are greater in number than those for L2 collocation research. Thus, different ways of measuring data from broadly similar instruments have been more fully explored. Second, since word association knowledge concerns the intuitive connections that an individual makes for the words in their mind, it can be understood as a more available type of connection than collocation knowledge. While an individual may well fail to produce conventionalised appropriate collocates for a particular lexical item, they may be more likely to produce associations if the lexical item is at least partially known. In that sense, word association can be thought to run ahead of collocation knowledge and may or may not interrelate with collocation knowledge as an individual develops control

of collocation connections.

Typically, in word association tests, single prompt words are used, and subjects are required to produce one or more responses which they associate with the prompt word. Responses are then often judged against adult native speaker norms. In L1 studies, it has been found that very young children tend to produce predominantly clang responses (i.e., responses that have a phonological rather than semantic link to the prompt word, as in **cat-bat**) and collocation responses (Aitchison, 1994, p.180). According to Aitchison (p.84), adult response types are characterised by greater semantic organisation, particularly with co-ordinates (e.g., **butterfly-moth, salt-pepper, left-right**). Collocation responses are the next most common type of response in adults, while superordination (e.g., **butterfly-insect**) and synonymy (e.g., **hungry-starved**) are less frequent. This change in response type between child and adult responses is generally referred to as the syntagmatic-paradigmatic shift.

Several L2 studies have explored whether such a shift also occurs with learners' word association responses as they become proficient (e.g., Söderman, 1993; Schmitt & Meara, 1997; Schmitt,1998b; Wolter, 2002). Learner responses are thus judged for their L1 adult native-ness, or stereotypy, but there are conflicting interpretations as to how consistent and pervasive the purported syntagmatic-paradigmatic shift is. Schmitt suggests that the shift is generally typical of both children and learners: "...the progression indicates the general organisation of lexical organisation patterns as a learner's language matures" (Schmitt, 2000, p.40). However, Söderman cautions against overgeneralisation and concludes that, with advanced learners, there is "a fairly stable balance between syntagmatic and paradigmatic responses when the stimulus words are frequent" (Söderman, 1993, p.168). Söderman's more conservative conclusion about L2 learners is remarkably similar to Aitchison's assertion that L1

adults produce both coordinate and collocation responses as the two most frequent types.

Not all L2 word association studies have been concerned with the syntagmatic-paradigmatic shift. Meara and Fitzpatrick (2000) approached handling word association responses by using a 30-item test in which the subjects were asked to write three or four responses to each stimulus. The items in their test were highly frequent (within the first 1000 most frequent words of English) and chosen for their characteristic of generating non-frequent responses. The data were first lemmatized, and the responses were lexically profiled for frequency, with one point given to each "infrequent" word, where infrequency was defined as outside the 1000 most frequent band. Meara and Fitzpatrick used the lexical profiling results to develop a "practical index of productive vocabulary" (p.26) for each individual.

In sum, the results from previous L2 word association research point to individual variation in responses, with some evidence of an increase in native-like associations at higher levels of proficiency. The findings also suggest that word associations are somewhat unstable and may show idiosyncratic, syntagmatic, or paradigmatic relationships to prompts. We have also seen that it is possible to use word association responses to sample productive vocabulary and develop individual lexical profiles based on the use of infrequent vocabulary.

7.2 Areas of focus

Using the results from previous studies to draw up some areas of focus for investigating the development of collocation knowledge over time suggests that instability may turn out to be a critical factor. Just as learners may not produce consistently the same word associations in response to a particular prompt word, they may also not produce the

same collocates over time. As word association studies have pointed to variation between individuals in the responses, we may also hypothesize that individual learners will differ in the collocates that they produce in response to the same prompt word. We may further theorize that individual learners will come up with a high number of varying non-native collocates, which may or may not become more native-like over time.

Although evidence from previous L2 collocation research for these suggestions is limited, we can note, for the design of the present study, a number of important points. First, the basic word association instrument of using a one-word prompt to elicit several target responses is time-efficient and simple to administer. Used on several occasions with a small group of words, such an instrument may produce a useful set of data for each individual. Second, as Meara and Fitzpatrick (2000) have shown, the results do not need to be scored for stereotypy or their native-like characteristics, but can be analysed in other ways to yield a particular lexical profile for each individual. Lastly, the same type of stimulus-response instrument may be efficiently used in a longitudinal study in two different ways. The instrument can be employed to elicit word associations over time so that we can develop a lexical profile of an individual learner. The same prompt words may also be exploited to produce collocates and provide evidence of the development of individuals' L2 collocation knowledge.

Table 7.1 below gives an example of the kind of data that such an instrument produces for the prompt word **advice**. There are four sub-categories of response. Word association responses are classified as non-syntagmatic or syntagmatic, and collocation responses are categorized as mis-collocations or appropriate collocations. The example learner produces just six different word associations, one of which, **doctor**, is also a collocation and therefore classified as a syntagmatic association. The

six word associations are all frequent, except for **professor**. The learner also produces just one mis-collocation (**refuse**) and generates seven appropriate collocates (**bad, give, good, ignore, reject, take,** and **useful**).

Table 7.1 Example word associations and collocations for ADVICE

Non-syntagmatic associations	parent, teacher, lawyer, friend, professor
Syntagmatic associations	doctor
Mis-collocations	refuse
Appropriate collocations	bad, good, ignore, reject, useful, give, take

In practice, the frequency of each response varies, and this information is presented later when the results of the study are reported. Nevertheless, the brief example in Table 7.1 shows that such an instrument is effective for eliciting both word associations and collocations.

7.3 Areas of exploration

In relation to the above-mentioned areas of focus, the following questions are explored in this study:

1. What is the lexical profile of the word associations that individuals produce over time?
2. Do individuals produce similar or different lexical profiles over time?
3. Do individuals vary their production of syntagmatic word associations over time?
4. What acceptable and unacceptable collocates do individuals produce over time?
5. Do individual acceptable and unacceptable collocate totals change over time?
6. Do individuals converge or diverge in the acceptable collocates that they produce over time?

7.4 The research instrument

A small corpus of texts on the theme of refugees and human rights was created, and 12 target nouns that were frequent in the corpus were selected for the experiment. The set of nouns was: **advice, assistance, asylum, attack, discrimination, homes, laws, persecution, protection, rights, rules**, and **war**. The set included both high frequency everyday nouns such as **advice, attack**, and **war**, as well as less frequent nouns with an academic profile. In each test administration, the subjects were required to write down three word associations for each of the 12 nouns, before they were asked to note three collocates that they believed would most often come together with each prompt noun. I decided to limit the target nouns to 12 because such a small number would yield two sets of a possible total of 36 responses per intervention; over six interventions, a maximum of 216 (6 x 36) word associations and 216 collocates could in theory be produced, if each response was distinct. The small set of nouns thus creates quite large individual corpora over time.

7.5 Subjects, procedure and scoring

The subjects for this study were 3 Law / Politics students—two male 3rd and 5th year students (hereafter referred to as Kensuke and Shoichi respectively) and one female 3rdyear undergraduate (identified as Emi). All three names are pseudonyms to ensure confidentiality. They had enrolled in a class on the theme of researching human rights issues through English, which they needed to pass for graduation credit. Refugees and human rights was an initial content focus in the course, after which students chose their own issues to research, read, discuss, and present on. The three students' previous experience of learning English had consisted mainly of reading and translation classes, though they each had taken English discussion courses earlier at university. They had

no prior experience of learning English collocations as an intrinsic part of their vocabulary development.

The word association and collocation tests were given during individual interviews held outside of class. The six administrations of the test occurred at the following points in the academic year: end of May (T1), mid-June (T2), end of June (T3), end of October (T4), start of December (T5), and mid-January (T6). This was not an ideal division of intervention times, and with a long summer break between the two semesters, it was impossible to space each intervention equidistantally. At the start of each test administration, the difference between word associations and collocation was carefully explained with examples. No feedback was given to the subjects about the nature of their word associations or the accuracy of their collocations. The order of items was randomly jumbled in each administration, and the answer sheet, once completed, was collected in.

The word associations were first lemmatized, following the same criteria as used by Meara and Fitzpatrick (2000), and then lexically profiled and categorized into four groups: K1, K2, AWL, and off-list. (K1 refers to the first one thousand most frequent words in English, and K2 to the next one thousand, while AWL covers items in the Academic Word List (Coxhead, 2000). Off-list includes all items outside those three wordlists.) A web-based automated program (Cobb, 2003) was used to obtain these profiles. Where more than three associations were written down, the first three were accepted for analysis. Proper nouns such as *America* and *UNHCR* were included in the data analysis; misspellings were corrected, and also accepted. Neologisms, however, were excluded from further analysis, while compound nouns such as *living place* and *human beings* were treated as off-list items. Each word association was given 1 point, and the totals for each of the four lexical categories were calculated for

each time. Following this, two scores for each subject were produced for T1-T6: The first score consisted of the total for K1 words (= high frequency associations); the second score came from the combined total for the other three categories (= infrequent associations). To treat all of the data sets as uniform sets of 36 responses, blank responses were counted as K1 words.

The students' collocations responses were examined by classifying them as acceptable or unacceptable. This was first done by reference to the *Oxford Collocations Dictionary for Students of English* (OUP, 2002). The dictionary is based on analysis of the British National Corpus (BNC) of 100 million running words, but does not indicate how frequent a particular collocate is or what its wider non-adjacent collocates are. To address this problem, the collocation sampler of the *Collins Wordbanks Online* English corpus (HarperCollins, 2004) was also used. The complete corpus consists of 56 million words of English, and 100 collocates for a search word can be retrieved together with their Mutual Information or *t* scores. The *t* score was used because, as Oakes notes (1998, p.194), "*t*-score collocates tend(ed) to have a higher frequency than the MI collocates." For example, the 10 most frequent *t* score collocates for **advice** are: **on, give, for, and, legal, information, citizens, bureau, seek**, and **help**. The corresponding top MI collocates are: **bureaux, nacab, capa, heeding, fatherly, bureau, impartial, CVs, citizens**, and **heeded**. Clearly, the *t*-score collocates are closer to what a learner might know. A collocation response was considered acceptable if it appeared either in the Oxford dictionary or in the list of *t* score collocates from *Collins Wordbanks*. In the case of **advice**, the Oxford dictionary on its own provided just over 50 collocates, but there were just over 100 possible collocates from using both sources (see Table 7.2 below). In short, using both sources allowed for a very broad range of acceptable collocates for each of the 12 nouns.

Table 7.2 **Database of appropriate collocates for ADVICE**

about	confidential	give	need	technical
accept	conflict	go to sb for	obtain	telephone
act on	constructive	good	offer	tip
advice	consumer	guidance	on	to
against	contact	health	or	train
also	confidential	heed	please	treatment
and	counsel	help	practical	turn to sb for
any	detail	helpful	professional	sensible
ask	doctor	how	provide	service
aspect	efficiency	if	receive	should
assistance	energy	ignore	reject	some
available	excellent	impartial	seek	useful
bad	expert	independent	sensible	valuable
best	financial	information	service	without
bit	follow	is	should	word
bureau	for	legal	some	wrong
can	free	listen to	sought	you
career	from	local	sound	your
centre	further	medical	specialist	
citizen	general	money	support	
clear	get	my	take	

The subjects' collocation responses were next classified into two groups, acceptable and unacceptable. Blank responses were treated as unacceptable collocates so that uniform data sets of 36 items could be maintained. Acceptable collocates (= hits) were given 1 point, as were unacceptable collocates (= misses), so that two scores (i.e., hits and misses) could be generated by subject for each prompt noun for T1-T6. The two scores were important for tracing changes in appropriate collocations and mis-collocations; a single score for each time period (i.e., total hits) was also calculated.

7.6 Word association results

The results for word associations are reported here and divided into four parts. In the first stage, the lexical profiles are reported, based on the index of frequent and infrequent associations for each intervention. Second, the total word associations for each individual are presented and then sub-divided into repeated and single associations. Finally, each individual's syntagmatic word association responses are presented, and an individual percentage is given as to how many syntagmatic responses have "shifted" by T6. These three sets of results allow us to form a quantitative picture of the three students' lexical profiles and word association behaviour. After we have looked at these results, we will consider the findings for collocations.

7.6.1 Lexical profiling of the word associations

The lexical profiles for the three students are presented in percentages in Table 7.3 below. Here, "Infrequent" covers K2, AWL, and Off-list items, i.e., excludes the 1000 most common words of English (K1).

Table 7.3 Infrequent lexical profiles of word associations

Subject	T1	T2	T3	T4	T5	T6
Emi	31%	47%	47%	50%	25%	42%
Kensuke	31%	53%	42%	53%	44%	67%
Shoichi	25%	58%	14%	19%	28%	33%

The results show that Emi and Kensuke have very similar profiles for T1-T5 and then begin to diverge in T6. Both Emi and Kensuke use frequent vocabulary and infrequent vocabulary almost equally at around 50%, before Kensuke starts to draw on infrequent

vocabulary more in T6. In contrast, Shoichi tends to use frequent vocabulary considerably more than infrequent vocabulary (see Figure 7.1 below) and is close to Emi in T5 and T6.

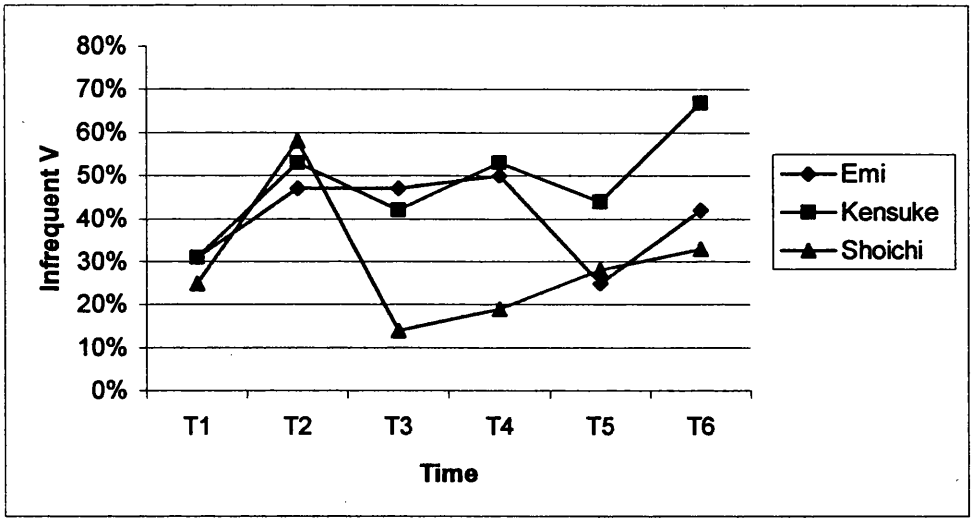

Figure 7.1 Lexical profiles of word associations over time

The overview of the word association results over time points to some temporary similarity between Emi and Kensuke; Shoichi has a somewhat different lexical profile from the other two, although his profile in T5 and T6 approaches that of Emi.

7.6.2 Total word associations

The total number of different word associations was estimated for each student for each of the 12 prompt nouns from T1 to T6 (see Table 7.4 below).

Table 7.4 **Total different word associations**

Noun	Emi	Kensuke	Shoichi
advice	6	12	12
rules	6	12	12
asylum	8	11	13
persecution	8	11	14
assistance	10	10	13
discrimination	10	13	13
homes	10	9	9
rights	10	7	13
laws	11	17	15
protection	12	12	12
attack	13	14	13
war	15	13	12
TOTAL	**119**	**141**	**151**

The table shows that Emi produced fewer associations in total than Kensuke or Shoichi. The results also indicate that Emi produced fewer different associations for each prompt noun than Kensuke (except in four cases: **homes**, **rights**, **protection**, and **attack**) and Shoichi (except in four cases, too: **homes**, **protection**, **attack**, and **war**). Kensuke produced an equal or greater number of associations than Shoichi, apart from three cases (**asylum**, **persecution**, and **assistance**).

The different word associations that each student produced were next sub-divided into two groups. The first set consisted of word associations that were repeated over T1-T6 more than once, and the second set comprised word associations that were produced just a single time. These two sets let us see how consistent each individual was in their word associations; the strength of this consistency provides a simple and partial indication of whether a particular individual tends to make many different

connections (i.e., may have a more loosely organized L2 lexicon) or to use fewer repeated connections (i.e., may have a more tightly organized L2 lexicon).

The findings are shown in Table 7.5 below. In the table, there are two columns for each individual. The first column shows the number of word associations that occurred more than twice. For **advice**, the entry for Emi is **5 (4, 4, 3, 3, 2)**. This means that Emi repeated 5 word associations over T1-T6. The distribution for these five repetitions is: 2 associations repeated 4 times, 2 three times, and 1 twice. The second column shows the frequency of single word associations. For Emi, the entry is **1 (1)**, indicating that she produced one single word association just once over the period of the study.

Table 7.5 Total number of different associations produced T1-T6

Noun	Emi		Kensuke		Shoichi	
advice	5 (4, 4, 3, 3, 2)	1 (1)	3 (3, 2, 2)	9 (1)	5 (3, 2, 2, 2, 2)	7 (1)
assistance	3 (5, 4, 2)	7 (1)	4 (4, 3, 2, 2)	6 (1)	1 (5)	11 (1)
asylum	3 (4, 3, 2)	5 (1)	3 (3, 3, 2)	8 (1)	2 (3, 2)	11 (1)
attack	3 (4, 2, 2)	10 (1)	4 (2, 2, 2, 2)	10 (1)	3 (3, 3, 2)	10 (1)
discrimination	4 (4, 4, 2, 2)	6 (1)	2 (3, 2)	11 (1)	4 (3, 3, 2, 2)	9 (1)
homes	3 (3, 2, 2)	7 (1)	3 (4, 3, 2)	6 (1)	3 (2, 2, 2)	6 (1)
laws	3 (4, 3, 2)	8 (1)	1 (2)	16 (1)	0	15 (1)
persecution	5 (4, 4, 2, 2, 2)	3 (1)	1 (2)	10 (1)	3 (4, 3, 2)	11 (1)
protection	4 (2, 2, 2, 2)	8 (1)	2 (4, 2)	10 (1)	5 (3, 2, 2, 2, 2)	7 (1)
rights	4 (5, 4, 3, 2)	6 (1)	5 (4, 3, 2, 2, 2)	2 (1)	3 (4, 3, 2)	10 (1)
rules	3 (5, 4, 3)	3 (1)	3 (3, 2, 2)	9 (1)	3 (3, 2, 2)	9 (1)
war	3 (2, 2, 2)	12 (1)	3 (2, 2, 2)	10 (1)	4 (3, 2, 2, 2)	8 (1)

For repeated associations, Emi has a mean score of 3.6, Kensuke 2.8, and Shoichi 3. For single associations, Emi tends to have lower totals—in only two cases (**attack** and

war) does she produce 10 or more associations, whereas Kensuke and Shoichi both generate 10 or more associations in 6 cases. The analysis shows a slight difference between Emi and the other two students. We may (very) tentatively conclude that Emi has a more tightly connected L2 lexicon—a point we return to later in the discussion.

7.6.3 Syntagmatic word associations

The results for syntagmatic shift are presented by individual for each of the 12 prompt nouns in Tables 7.6 – 7.8 below. In these tables, each row shows the syntagmatic response word association responses. The syntagmatic responses are presented in plain style for their first appearance and for each successive appearance, except for their final appearance before T6. If a previously occurring syntagmatic word association response re-occurs by T5 but does not appear in T6, such a response is highlighted in ***bold italics*** on its final appearance. In some cases, syntagmatic responses occur once and once only, in which case they have been also put in bold italics. It is these ***bold italicised*** responses that are counted as syntagmatic shifts in the brief commentary following each of the three tables. Here, total syntagmatic word association responses for each individual are given, together with a percentage figure for the same responses which were not re-produced by T6. This percentage lets us establish a sense of shift for each individual.

Table 7.6 **Emi's syntagmatic word association responses**

Noun	T1	T2	T3	T4	T5	T6
advice	doctor	doctor	--	--	***doctor***	--
assistance	***finance***	--	--	--	***help***	--
asylum	--	refugee	refugee	refugee	***governmen t***	refugee
attack	***terrorist***	***physical, verbal***	--	--	--	----
discrimination	***race, sex***	--	--	--	***rights***	--
homes	--	house	family, house	***family***	--	house
laws	--	***protection***	--	--	***court, rule***	--
persecution	--	--	--	--	--	--
protection	law	***child, law***	--	--	--	--
rights	human	***child,*** human	***freedom,*** human	human, ***woman***	***law***	human
rules	--	law	game, law	game, law	law	game, law
war	***refugee***	***country***	--	***criminal, Iraq***	--	--

Table 7.6 shows that Emi produced a total of 30 syntagmatic responses, of which 24 (80%) shifted by T6. The evidence for syntagmatic shift with Kensuke is presented in the next table below.

Table 7.7 Kensuke's syntagmatic word association responses

Noun	T1	T2	T3	T4	T5	T6
advice	help	--	help	useful	help	helpful, useful
assistance	food, help	food	help	food	***help***	aid, food
asylum	--	--	--	refugee	refugee	refugee
attack	--	***bomb***	--	--	--	--
discrimination	--	--	--	--	--	--
homes	--	--	country	***country***	--	--
laws	***regulation***	***strict***	--	***control, rights***	***justice***	constitution, criminal, international
persecution	--	--	--	--	--	--
protection	--	--	--	--	--	--
rights	--	law	***law***	human, legal	***human,*** legal	legal
rules	law	***law***	***absolute, strict***	***break, formal***	--	--
war	***refugee***	army	--	***terrorism***	***army***	--

Table 7.7 indicates that Kensuke produced a total of 27 different syntagmatic responses, of which 17 (63%) made the shift by T6.

In Table 7.8 below, the results for Shoichi show that he generated a total of 29 different syntagmatic responses, of which 20 (69%) shifted by T6.

Table 7.8 Shoichi's syntagmatic word association responses

Noun	T1	T2	T3	T4	T5	T6
advice	--	help	help	--	helpful	--
assistance	help	***food***	help	help	help	help
asylum	--	--	--	--	--	--
attack	army	--	***army***	***force***	--	--
discrimination	race	race	--	***race. woman***	--	black
homes	--	--	***building***	family, house	***house***	family
laws	--	--	--	--	***court***	rights
persecution	--	victim	victim	--	***victim***	country, German
protection	--	--	animal, law	***environ- ment, law***	***animal***	--
rights	human, ***obligation***	human	***freedom***	***human,*** law	***law, lawyer***	citizen
rules	--	--	--	game	--	game
war	***country***	--	***fight,*** world	bad	***bad,*** world	world

Overall, the raw totals of syntagmatic responses are very close for the three students (Emi: 30, Kensuke: 27, Shoichi: 29), but the percentage shift differs (Emi: 80%, Kensuke: 63%, Shoichi: 69%). However, it cannot be claimed that the items which have "shifted" by T6 necessarily indicate a corresponding increase in paradigmatic responses for each individual. The observable data simply do not allow us to infer this.

7.7 Collocation results

The collocation results are based on the three collocates that the subjects wrote down in

each test administration for the 12 target words. The results are reported in six stages. First, collocation profiles are presented, based on the percentage totals of acceptable and unacceptable collocations for each intervention. Second, the total different collocations are given for each individual, and the distribution of different appropriate collocations for each intervention is reported. The third stage of the analysis focuses on the total distribution of all appropriate collocations produced by each individual. We then look at the total of mis-collocations for the three students, before, in the following stage, we see whether the students produce the same collocates or not. Finally, results for appropriate collocate types and frequencies are presented for each of the trio. These six sets of results let us develop an overview of the longitudinal development of the students' collocation knowledge, as well as identify quantitative and qualitative variations.

7.7.1 Collocation profiles

To develop a collocation profile of each individual, the percentage totals of acceptable collocations (referred to as "hits") and unacceptable collocations (referred to as "misses") were estimated for T1-T6. Table 7.9 below shows these results.

Table 7.9 Appropriate collocation profiles

Subject	T1	T2	T3	T4	T5	T6
Emi	31%	56%	61%	67%	64%	72%
Kensuke	25%	42%	47%	39%	47%	47%
Shoichi	36%	11%	33%	17%	56%	33%

Figure 7.2 below translates these percentages into a graph with the collocation profiles of the three students for T1-T6.

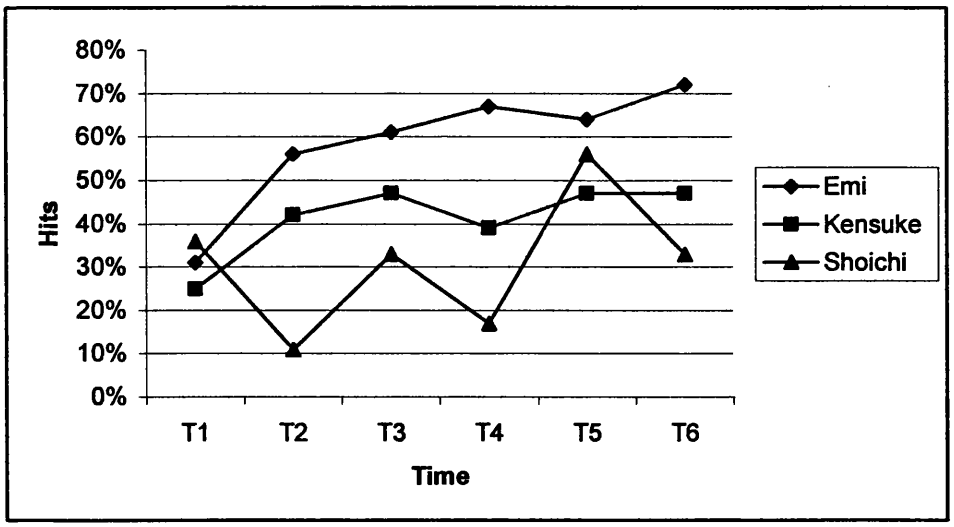

Figure 7.2 Collocation profiles over time

Overall, the results tend to indicate that Emi has a distinctly different collocation development from Kensuke and Shoichi. According to Figure 7.2, after T1, Emi produces increasingly more acceptable collocates, except in T5. Kensuke starts by increasing his "hit" rate, but from T2 on there is little variation in his performance. On the other hand, Shoichi fluctuates over the academic year, and his score at T1 is very close to his scores at T3 and T6. In brief, we seem to have three distinct collocation profiles.

7.7.2 Total different appropriate collocations

The total different appropriate collocations were estimated for each individual. The results are shown in Table 7.10 below. It would be theoretically possible for each learner to score a maximum of 18, were each learner to produce three different appropriate collocates for each stimulus noun over six interventions.

Table 7.10 Total different appropriate collocations

Noun	Emi	Kensuke	Shoichi
advice	7	9	8
assistance	5	7	4
asylum	2	3	4
attack	4	5	1
discrimination	5	5	7
homes	2	2	1
laws	5	3	3
persecution	1	1	1
protection	7	5	4
rights	6	6	3
rules	4	3	3
war	4	4	3
TOTAL	**52**	**53**	**42**

Emi and Kensuke have nearly identical scores, while Shoichi produces approximately 20% fewer appropriate collocations.

7.7.3 Repetition of different appropriate collocations

The repetition of different appropriate collocations was examined to understand better the patterns in the collocation profiles of the three students. The results are shown in Table 7.11 below. In the table, the first figure in the column for each individual shows the number of times that an appropriate different collocation was produced. After this, the figures in brackets show the frequency of each different appropriate collocation. For **asylum**, the entry for Emi is **2 (6, 1)**. This means that Emi produced two appropriate collocates—the first 6 times, and the second once. In contrast, the entry for Kensuke is **3 (4, 3, 1)**. Kensuke generated three different acceptable collocates for

asylum—the first 4 times, the second 3 times, and the third once. Shoichi produced four different appropriate collocates just one time each, so his entry is **4 (1, 1, 1, 1)**.

Table 7.11 Distribution of different appropriate collocations

Noun	Emi	Kensuke	Shoichi
persecution	1 (1)	1 (1)	1 (2)
homes	2 (5, 1)	2 (1, 1)	1 (3)
asylum	2 (6, 1)	3 (4, 3, 1)	4 (1, 1, 1, 1)
rules	4 (5, 3, 3, 1)	3 (5, 4, 1)	3 (4, 1, 1)
attack	4 (5, 4, 1, 1)	5 (3, 1, 1, 1, 1)	1 (1)
war	4 (5, 4, 4, 1)	4 (4, 1, 1, 1)	3 (4, 2, 1)
discrimination	5 (3, 2, 2, 1, 1)	5 (3, 1, 1, 1, 1)	7 (3, 2, 2, 1, 1, 1, 1)
assistance	5 (4, 3, 2, 2, 1)	7 (3, 1, 1, 1, 1, 1, 1)	4 (2, 1, 1, 1)
laws	5 (6, 2, 2, 1, 1)	3 (2, 2, 1)	3 (5, 2, 1)
rights	6 (5, 4, 2, 1, 1, 1)	6 (4, 3, 2, 2, 1, 1)	3 (3, 2, 1)
protection	7 (4, 3, 2, 2, 1, 1, 1)	5 (3, 2, 1, 1, 1)	4 (3, 1, 1, 1)
advice	7 (5, 4, 1, 1, 1, 1, 1)	9 (5, 1, 1, 1, 1, 1, 1, 1, 1)	8 (5, 2, 1, 1, 1, 1, 1, 1)

The results point to differences between Emi and the other two students. In 10 cases, Emi reproduces the same collocate four times or more, the exceptions being **discrimination** and **persecution**. In seven of those 10 cases, she also manages to repeat a second collocate three times or more. On the other hand, Kensuke manages to reproduce a second collocate three times or more in just three cases (**asylum**, **rights**, and **rules**). In six cases, Kensuke produces second and other collocates just once (**advice**, **assistance**, **attack**, **discrimination**, **homes**, and **war**). Shoichi produces a second appropriate collocate twice in five cases (**advice**, **discrimination**, **laws**, **rights**, and **war**); in five other cases, Shoichi supplies second and other collocates just once (**assistance**, **asylum**, **protection**, **rights**, and **rules**). What these findings appear to suggest is stronger collocation links for Emi in that she tends to repeat

different collocates more times than the other two students. While both Kensuke and Shoichi do this to some extent, their collocation production is marked by weaker reproduction of other, secondary collocates. Emi has strength of repetition in the collocates that she produces; Kensuke and Shoichi both have a wide range of collocates, but many of these occur just once. In a word, their collocation knowledge seems thinly spread.

7.7.4 Mis-collocations

Mis-collocations were totalled for each of the three students, as shown in Table 7.12.

Table 7.12 **Mis-collocation totals**

Noun	E	K	S
advice	1	2	9
protection	1	7	4
rights	1	3	6
laws	2	7	3
rules	2	5	4
war	2	6	3
attack	3	4	9
homes	3	6	6
assistance	4	4	6
asylum	4	6	8
discrimination	4	6	6
persecution	7	7	7
TOTAL	**34**	**63**	**71**

Emi has roughly 50% fewer mis-collocations than either Kensuke or Shoichi, whose mis-collocation totals are rather close. Emi varies from the other two students in the accuracy of her collocation knowledge.

7.7.5 Overlaps in appropriate collocations

The penultimate stage of the quantitative collocation analysis examined how many identical collocates were produced by the three students (see Table 7.13 below).

Table 7.13 Overlaps in appropriate collocations

Noun	Collocate
advice	GIVE, USEFUL
assistance	GET, GIVE, OFFER
asylum	SEEK(ER)
attack	--
discrimination	--
homes	--
laws	--
persecution	--
protection	GIVE, OFFER
rights	HUMAN, PROTECT
rules	MAKE, BREAK
war	--

The students produced 147 different appropriate collocates (as shown in Table 7.10), but, according to Table 7.13, in only 12 cases did the trio generate the same collocates. The individual variation for collocation knowledge is, in other words, fairly comprehensive.

7.7.6 Appropriate collocate types and frequencies

The final part of the collocation results features the different appropriate collocates that the three students produced. These are shown by individual in Tables 7.14-7.16 below.

Table 7.14 Acceptable collocates produced by Emi T1-T6

Noun	Acceptable collocations	Total
advice	TAKE (5), GIVE (4), BAD (1), GOOD (1), IGNORE (1), REJECT (1), USEFUL (1)	**14**
protection	GIVE (4), GET (3), FINANCIAL (2), LEGAL (2), HAVE (1), HUMAN RIGHTS (1), OFFER (1)	**14**
rights	HUMAN (5), PROTECT (4), BASIC (2), CHILDREN (1), RESPECT (1), VIOLATION (1)	**14**
war	GO TO (5), CIVIL (4), COLD (4), CRIMINAL (1)	**14**
assistance	GET (4), FINANCIAL (3), GIVE (2), OFFER (2), GOVERNMENT (1)	**12**
laws	ENACT (6), BREAK (2), PASS (2), OBEY (1), STRICT (1)	**12**
rules	MAKE (5), BREAK (3), OBEY (3), STRICT (1)	**12**
attack	PHYSICAL (5), VERBAL (4), VIOLENT (1), SUDDEN (1)	**11**
discrimination	SUFFER (3), MAKE (2), SEXUAL (2), BETWEEN (1), UNFAIR (1)	**9**
asylum	SEEK(ER) (6), LEAVE (1)	**7**
homes	LEAVE (5), FLEE (1)	**6**
persecution	AGAINST (1)	**1**

Table 7.15 Acceptable collocates produced by Kensuke T1-T6

Noun	Acceptable collocations	Total
advice	ASK (1), GIVE (5), ACCEPT (1), GOOD (1), HELPFUL (1), NEED (1), OFFER (1), PROVIDE (1), USEFUL (1)	**13**
rights	GET (4), LOSE (3), HAVE (2), HUMAN (2), LEGAL (1), PROTECT (1)	**13**
rules	MAKE (5), BREAK (4), STRICT (1)	**10**
assistance	ASK (1), PROVIDE (3), FINANCIAL (1), GET (1), GIVE (1), OFFER (1), RECEIVE (1)	**9**
asylum	PROVIDE (4), SEEK(ER) (3), REFUGEE (1)	**8**
protection	LEGAL (3), PROVIDE (2), GIVE (1), NEED (1), OFFER (1)	**8**
discrimination	SUFFER (3), ELIMINATE (1), RACIAL (1), STOP (1), UNDER (1)	**7**
war	NUCLEAR (4), COLD (1), PREVENT (1), WORLD (1)	**7**
attack	HEART (2), ENEMY (1), HAVE (1), NUCLEAR (1), SUDDEN (1)	**6**
laws	CRIMINAL (2), INTERNATIONAL (2), STATE (1)	**5**
homes	LOSE (1), THEIR (1)	**2**
persecution	RACIAL (1)	**1**

Table 7.16 Acceptable collocates produced by Shoichi T1-T6

Item	Acceptable collocations	Total
advice	GIVE (5), ASK (2), GET (1), OFFER (1), PROVIDE (1), RECEIVE (1), TAKE (1), USEFUL (1)	**13**
discrimination	BLACK (1), RACIAL (2), STOP (2), AGAINST (1), MAKE (1), PROHIBIT (1)	**8**
laws	ENACT (5), BREAK (2), BY (1)	**8**
war	START (4), WORLD (2), BEGIN (1)	**7**
protection	PROVIDE (3), FROM (1), GIVE (1), OFFER (1)	**6**
rights	HUMAN (3), PROTECT (2), ANIMAL (1)	**6**
rules	MAKE (4), BREAK (1), ESTABLISH (1)	**6**
assistance	PROVIDE (2), GET (1), GIVE (1), OFFER (1)	**5**
asylum	GIVE (1), HAVE (1), PROVIDE (1), SEEK (1)	**4**
homes	BUILD (3)	**3**
persecution	SEVERE (2)	**2**
attack	HAVE (1)	**1**

In the previous sub-section, we established that there is very little overlap between the three students in the types of collocates that they produced. However, if we lexically profile the collocates in Tables 7.14-7.16, we can see that Emi and Kensuke draw on frequent and infrequent vocabulary to a quite similar degree, whereas Shoichi has a slightly different profile (see Table 7.17).

Table 7.17 Lexical profile of acceptable collocates

Collocates	Emi	Kensuke	Shoichi
Frequent	74%	79%	88%
Infrequent	26%	21%	12%

Despite this similarity, Emi has a greater range of infrequent collocates than Kensuke, who in turn has a bigger variety than Shoichi (see Table 7.18).

Table 7.18 Range of infrequent acceptable collocates

Subject	Infrequent collocates
Emi	civil, enact, financial, flee, ignore, legal, physical, reject, seek, seeker, sexual, verbal, violation
Kensuke	eliminate, financial, legal, nuclear, racial, refugee, seek, seeker
Shoichi	enact, establish, prohibit, racial, seek

However, Emi and Kensuke show similarity in the grammatical classes that they use for acceptable collocates, as indicated in Table 7.19.

Table 7.19 Grammatical profile of all acceptable collocates

Class	Emi	Kensuke	Shoichi
Verbal	50%	57%	76%
Adjectival	35%	31%	12%
Nominal	11%	10%	5%
Prepositional	4%	5%	7%

All three students use verbal collocates the most, but Emi and Kensuke use adjectival combinations much more than Shoichi.

7.8 Discussion

It is perhaps good to start the discussion by underlining the benefit of exploring in such extensive detail the word associations and collocates of three students with a small set of prompt nouns. This has allowed us to hold in view the particular productions of individuals so that we do not lose sense of the striking variation between individuals. There are, of course, questions about how far we can generalize from such data, but

within the overall individual variation in both word association and collocation knowledge between the three students, there are several intriguing differences that deserve our attention.

The use of frequent and infrequent vocabulary is the first. The lexical profiling of the word associations showed that Shoichi uses frequent vocabulary relatively more often than either Kensuke or Emi. Similarly, Shoichi was found to use frequent vocabulary more for acceptable collocations and to have a smaller range of infrequent collocates than the other two students. Shoichi also had the lowest total of different appropriate collocates and the highest number of mis-collocations. We further observed that Shoichi had a fluctuating collocation profile in terms of hits and misses. Can we infer any connection between all these different observable results? It is not completely clear, but it appears that Shoichi fluctuates because he has a smaller productive vocabulary (see Figure 7.1) than Kensuke or Emi and is at the limit of his productive capacity for achieving collocation appropriacy. He necessarily activates frequent vocabulary more often for associations and collocations.

This interpretation tends to suggest that Shoichi is the least collocationally competent of the three students, but it is not altogether obvious whether this is in fact the case. To shed further light on this question, it is useful to compare each subject's total collocation "hits" (i.e., acceptable collocates) for the first half of the study (T1-T3) with their total scores for the second half (T4-T6). If an individual scores higher for the second half, we may take this as a general indicator of increasing collocation development, while the opposite scenario would suggest stagnating or faltering development (see Figure 7.3 below).

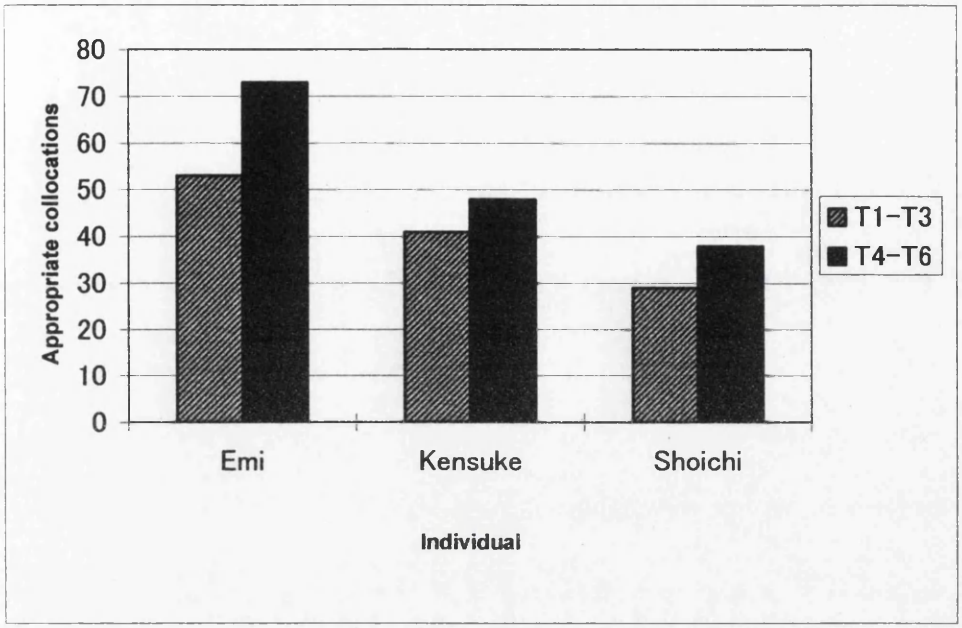

Figure 7.3 Comparison of collocation profiles T1-T3 and T4-T6

Figure 7.3 indicates that all three individuals score higher in the second half, but Emi and Shoichi show greater difference between the two periods than Kensuke does. From this point of view, Shoichi makes clear progress; his overall fluctuating performance seems then to indicate lexical growth spurts.

The use of frequent and infrequent vocabulary by Emi stands in contrast to Shoichi. The lexical profiling of word associations showed that Emi tends to draw on both types of vocabulary somewhat equally, or with something approaching a 60-40 balance in favour of frequent vocabulary. At the same time, she has a wider range of lexically infrequent appropriate collocates than Shoichi. This seems to mark one difference in collocation knowledge, as if she has a larger vocabulary in the first place. Other results allow us to describe this difference more closely. First, compared to the other two students, Emi shows a constant positive increase in total appropriate collocates.

The second characteristic is that Emi produces fewer different associations and mis-collocations, while she also has greater repetition of more than 1 acceptable collocate in nearly all cases. We suggested for the word association results that Emi tends to have a more tightly organised L2 lexicon; it appears, too, that she has fewer collocation links and that she can repeatedly activate these links. It is as if she doesn't need to search between competing options in her L2 lexicon for either word associations or collocates, but can access the same items consistently over time. This interpretation suggests that her lexical and collocation links are stronger and relatively more available for production.

On the other hand, Kensuke remains in the middle between Emi and Shoichi. His lexical profile for word associations is closer to Emi, and he produces an almost identical total of different acceptable collocations as she does. Yet, the collocation profile showed that Kensuke flat-lines in his collocate production over time, making no obvious progress between T2 and T6. One plausible explanation is that he has a large vocabulary like Emi, but his lexicon is more loosely organised for both word associations and collocations. Kensuke can make many different links for both types of knowledge, but cannot repeat the same links as relatively often as Emi. His lexical and collocation knowledge appears, as we noted earlier, to be thinly spread, so he produces more single word associations and collocations than Emi.

The differences between the three learners seem to point to some rather elusive variations in the way that their L2 mental lexicons are individually organised. There are, of course, problems in generalizing from results based on such a small set of prompt items, and such limitations are discussed in more detail below. At the same time, we know so little about L2 collocation development that it is worth temporarily looking in detail at the small picture in order to question what the larger canvas of L2

lexical organisation might involve. Indeed, the starting concern of this study was to develop a research tool that would allow us to examine collocation development over time. We did not want to look at collocation development in isolation; rather, we aimed to examine word association knowledge as well, so that we could see whether any interrelationships could be found between the two types of knowledge for these three learners. We noted how previous L2 collocation studies have tended to involve single interventions and not looked at collocation development longitudinally. One weakness of such synchronic research is to construe the development of collocation knowledge in negative terms by focusing on errors in collocation production, separately from learner success. Hopefully, the present study redresses the balance somewhat by showing that rates of acceptable and unacceptable collocation production are best viewed together. They have a dynamic and variable interrelationship for different individuals. This study has also shown that it is possible to elicit quite a large number of collocates for an individual in a relatively simple and effective way. Furthermore, there are potentially important connections to observe between frequent and infrequent vocabulary, word association production, and collocation production. Chief among these hypothesized connections is the notion that a similar range of productive vocabulary does not necessarily entail similarity in productive collocation ability. It appears that it is rather the varying activation between links in L2 lexicons of broadly similar size that is critical (cf. Meara & Wolter, 2004).

7.9 Limitations

The study suffers from several limitations which are worth comment. As mentioned above, the first weakness is the limited number of items. The instrument used just 12 words, two of which did not function particularly well for collocation

responses—**homes** and **persecution**. The plural form of **homes** is relatively rare, and so it was difficult for students to make appropriate connections; **persecution** proved to be an unknown lexical item for all three students at the beginning. Although the students managed to produce word associations for this noun later, the number of acceptable collocates produced by each individual by T6 was very low. For these reasons, it would be wise to include a greater number of more frequent items if the research instrument were to be further developed. By using more frequent items as the stimulus words, it would be possible to get a broader sampling of the L2 lexicon in terms of lexical profiling; we would also get a broader sense of collocation production. A further weakness of this study is that it may suffer from a test-re-test effect. It may be better in future either to reduce the number of interventions or to use a separate set of prompt words for the word associations and the collocate elicitation. A final limitation is that basic quantitative analysis alone does not explain how learners go about the challenge of developing their collocation ability in English. We have derived some very broad indicators such as the lexical growth spurts for Shoichi, flat-lining for Kensuke, and overall positive collocation development of Emi, but we have not been able to interpret this further in this study.

7.10 Conclusion

We have, in this chapter, examined a small-scale longitudinal study of individual variation in L2 word association and collocation knowledge. The study was based on adapting a word association measure to see how well it can give insights into both the lexical profiles of individual learners and the development of their collocation knowledge over one academic year. The differences between the three subjects suggested that we need to pay much greater attention to individual variation than has

previously been attempted in L2 collocation research. To understand such individual variation qualitatively, it would be useful to interview learners over time too and develop a dialogue with them about how (and why) they go about organizing their vocabulary and collocation learning. We will return to this concern in Chapter 9, but, before that, we will turn our attention to adapting and improving the research instrument for a larger scale quantitative investigation with a bigger population. This is the focus of the next chapter.

Chapter 8

Developing a Productive Test of Collocation Knowledge

8.0 Overview

The longitudinal case study in Chapter 7 was restricted to a small number of 12 prompt items. Despite providing a number of interesting insights, the interpretation of the data remained one of speculative plausibility rather than firm conclusion because we could not for certain prove whether the results were random or statistically significant. That weakness was in the nature of the beast. We wanted to look closely at the development of collocation knowledge over time and keep our eyes on the actual word associations and collocates that the three students produced, rather than limit the focus to quantitative data alone. On the plus side, the adaptation of the stimulus-response instrument proved useful for eliciting both word associations and collocates; moreover, the careful analysis of individual data over time suggested that the instrument could be developed into a useful tool for exploring collocation knowledge in a more statistically rigorous fashion with a larger population. In this chapter, we will pursue that goal.

8.1 Introduction

One of the key questions that vocabulary testing raises is what items should be included in a particular instrument. In the tests presented in previous chapters, both highly frequent items and infrequent items have been used; the basic underlying assumption has been that both types of items are useful for tracking and observing differences in collocation knowledge between less proficient and more advanced learners of English. However, the results presented in Chapter 7 suggest that highly frequent vocabulary items only should now be used for expanding the number of nouns used for testing collocation knowledge. We noted, for example, in the previous chapter that infrequent

items such as **persecution** did not function well and that the three students struggled to supply appropriate collocates for this noun over the 9-month period of the study. In contrast, frequent items such **advice**, **attack**, and **war** performed well, providing a range of appropriate collocation responses. An exclusive focus on frequent items may therefore allow us to see how both less proficient and more proficient learners collocate lexis that they can confidently recognize. We inferred that the range of collocation responses pointed to individual differences in the organization of the L2 lexicon. Though the students seemed to have English vocabularies of roughly similar size, they tended to select and re-activate different collocation links for frequent vocabulary. Using a greater number of frequent nouns in an expanded test may well allow us to understand further differences that underlie L2 collocation production.

The inquiry in this chapter will focus on using a stimulus-response instrument with a larger set of highly frequent noun prompts. We will first explore the piloting of 50 nouns with both native speakers and highly proficient non-native speakers in order to make an informed selection of 30 items. We will then investigate how such a 30-item test works with a group of learners who range in English proficiency from low intermediate to advanced.

8.2 Research focus

This study explores whether there is a difference between learners of low and high English proficiency in their production of frequent and infrequent collocates.

8.3 Method

A 30-item stimulus-response test was designed to elicit three collocates for each item. The test was derived from the instrument reported in longitudinal case study reported in the previous study, but contained new, highly frequent items that were selected after

piloting with both native and non-native speakers of English. The selected items were further normed with a large group of English native speakers to provide independent baseline data.

8.3.1 Selection and norming of the target items

Fifty highly frequent nouns were chosen from a lemmatized list of the 500 most frequent items in the British National Corpus (Kilgarriff, 1996). The selection of nouns is shown in Table 8.1 below.

Table 8.1 50 nouns selected for piloting

back	example	home	money	research
body	experience	house	paper	role
book	eye	idea	patient	sense
car	face	interest	plan	support
child	family	issue	point	table
country	friend	law	police	value
death	government	letter	power	voice
decision	hand	life	problem	war
door	head	line	question	water
end	health	mind	reason	work

The 50 nouns were piloted with 2 groups—one, a highly proficient group of 35 Japanese users of English, and the other, a group of 35 British native speakers. In this pilot, the two groups were asked to provide three verbs that they would most expect to come before each noun. Given that such frequent nouns often have over 100 possible collocates, I decided to restrict the elicitation to transitive verbs only, so that I could secure a small set of directly comparable responses from either group. Each verb was scored as 1, whether it was the primary response or not, and the verb collocate responses were then compared between the two groups. Table 8.2 below shows the results for two example items. For the first item, **door**, both the NNS and NS groups produced a limited set of almost identical collocates, which suggested that this item

would not function well for the main test. In contrast, the second item, **work**, led to a wider range of divergent collocate responses between the two groups.

Table 8.2 Example non-divergent and divergent items

Non-divergent item		
DOOR	**NNS**	**Open** 35, **Close** 33, **Shut** 19, **Knock** 11 *Others*:* **Break, Lock, Slam**
	NS	**Open** 31, **Close** 30, **Shut** 15, **Slam** 7, **Lock** 6
Divergent item		
WORK	**NNS**	**Go** 15, **Have** 14, **Do** 11, **Start** 8 *Others:* **Engage, Find, Give, Hate, Keep, Leave, Like, Look, Quit**
	NS	**Do** 16, **Go** 15, **Enjoy** 8 *Others:* **Find, Finish, Get, Have, Leave, Look, Make, Seek, Start**

* Others = Items which produced five or fewer responses

Thirty nouns producing mainly divergent results between the two groups were then chosen for use in the collocation test. These nouns are presented in Table 8.3 below (see Appendix 8.1 for detailed results from this initial piloting and selection).

Table 8.3 30 nouns selected from NS & NNS piloting

body	example	house	police	role
car	experience	interest	power	support
child	family	issue	problem	value
country	friend	law	question	voice
death	government	life	reason	war
decision	health	paper	research	work

To establish a database of possible collocates for each target noun, two sources were used: *Collins Wordbanks Online* (HarperCollins, 2004) and the *Oxford Collocations Dictionary* (OUP, 2002). The sources were combined to create a set of collocates for each noun, and the 30 collocate sets were then lexically profiled and sub-divided into frequent and infrequent collocates. As in Chapter 7, frequent collocates

consisted of those items within the 2000 most frequent words of English, whereas infrequent collocates involved Academic Word List as well as off-list items (see Appendix 8.2 for the raw totals of frequent and infrequent collocates for each item). These collocates were used for scoring the results from the test; the 30 nouns were also normed with a group of 50 native speakers of English resident in Japan (see Table 8.4 below for a bio-data summary of the native speaker population).

Table 8.4 Native speaker norming bio-data summary

Descriptor	Result
Female	19
Male	31
Mean age	42.5
Minimum age	28
Maximum age	60
Mean years resident in Japan	13.5
Minimum years resident in Japan	1
Maximum years resident in Japan	35
Nationality: *American*	24
British	12
Australian	7
Canadian	6
New Zealand	1

The native speaker norming was carried out in order to explore similarities with non-native speaker performance in the main test administration, as well as with the combined corpus-derived sets described above.

8.3.2 Test procedure and scoring

Two different pencil-and-paper forms of the 30-item test were prepared, with items randomly jumbled in either form for counterbalancing. After a brief explanation and guided practice with three additional items (**holiday**, **letter** and **university**), the test-takers were asked to write down three collocates for each of the 30 nouns. They had 30

seconds for each noun. The test administration lasted approximately 20 minutes total. The students belonged to different 1st, 2nd and 3rd year undergraduate Faculty of Law English classes, ranging in proficiency from low intermediate through to advanced. Their TOEIC scores from the institutional placement test at the start of the academic year were used as independent indicator of their overall English ability.

The data were entered into the computer, with misspellings corrected. In the few cases where multi-word combinations occurred, they were reduced to the main lexical element (for example, **take care of** became **care**), and phrasal verbs were scored for the base verb (for instance, **come** for **come up with**). Uninflected base verb responses were accepted where the database indicated participle forms only. For example, **lose** and **love** were scored as appropriate collocates for **friend**, although the database included only **lost** and **loved**, and **help** was accepted for **police** (database: **helping**). These minor adjustments were made so that subject responses could be treated at a lexical rather than grammatical level for appropriacy.

The subjects' responses were scored as 1 or 0 according to whether they matched the collocate database. Blank responses were also scored as 0 to create intact data sets. This first score produced a total for appropriate collocate responses, with a maximum of 90 possible. Two groups of 20, one high and the other low, were formed by TOEIC score. The students with the 20 highest and lowest TOEIC scores were used to form the high and low groups. The low group's TOEIC scores ranged from 325 to 430 with a mean of 389, and the high group's from 680 to 900 (mean: 787). In the next stage of the analysis, each individual's collocate responses were lexically profiled and categorised as frequent or infrequent, with blank responses treated as frequent in order to maintain intact data sets. Appropriate collocate responses were then separated into frequent and

infrequent, and two further individual scores—appropriate frequent collocate responses and appropriate infrequent collocate responses—were derived for each of the 40 subjects.

Table 8.5 below shows two examples of how the scoring applied to two subjects' responses for the test item **decision**. Both subjects produced three collocates each. All three responses from A were frequent, of which only two were appropriate (= hits): **important** and **make**. Two of Subject B's responses were frequent and one infrequent (**final**), and all three were appropriate.

Table 8.5 Example responses and scoring

Subject	Item	Collocates	Frequent Collocates	Infrequent Collocates	Frequent Hits	Infrequent Hits
A	decision	important make my	important make my	--	important make	--
B	decision	final make reach	make reach	final	make reach	final

8.3.3 Analytical approach

I report the findings from the complete analyses in more detail below, but, before we look at the results, I would like to comment briefly on the method of analysis. I mentioned above that I decided to compare the results between two groups of learners, one high proficiency and the other low proficiency. I took this decision in the expectation of deriving some significant results between the two groups. There is, in fact, nothing unusual about this, since more proficient learners will routinely score better than less proficient learners on measures of their English ability. Such an approach is useful in that it helps us understand how groups of different ability cope

with the same set of items: significant differences between groups let us get a sense of development between less proficient and more proficient learners. However, it is important to emphasize that group-based analysis takes us away from understanding individual learner performance and from developing a predictive model of learner ability and item difficulty. McNamara notes in a discussion of classical statistical analysis that "...we have no way of knowing whether these characteristics of person ability and item difficulty would be maintained for the persons over different items if they were tried out on different subjects" (McNamara, 1996, p.153). So, although we may get some good results from group-based analysis, the findings may tend to be somewhat restricted and routine.

This raises the question as to what kind of analysis can help us balance the focus on generalized difference between groups with a concern for understanding and predicting individual performance. The answer lies, I believe, in item response theory (IRT), where the Rasch model can enable us to examine within a given population the "probabilistic relation between any item's difficulty and any person's ability" (Bond & Fox, 2001, p.199). Detailed explanations of the Rasch model are available in McNamara (1996) and Bond and Fox (2001), so I will summarize here briefly the main features of the approach to provide a clear picture of this alternative form of analysis.

Rasch measurement works by estimating subjects' performance on a test and calculating "from the data the chances of a candidate of a given ability achieving a certain score on an item of a given difficulty" (McNamara, 1996, p.152). It achieves this extrapolation by reiteratively recalculating the match between the predicted score set and the observed score set "until the required level of accuracy is met" (McNamara, 1996, p.162). This calibration is known as "maximum likelihood estimation"

(McNamara, 1996, p.161) and produces a logit scale to represent the relationship between person ability and item difficulty. Item difficulty is conventionally set at 0 logits on the logit scale, so that items of "above-average difficulty will be positive in sign, those of below-average difficulty negative in sign" (McNamara, 1996, p.165-6). Rasch analysis also generates an item ability map to give a graphic display of the fit between person ability and item difficulty.

The fit statistics generated in Rasch analysis by the computer program WINSTEPS (Linacre, 2005) are comprised of infit mean square, outfit mean square, infit *t*, and outfit *t*. Fit here refers to how well the response pattern for individual items matches the general pattern of response generated by the calibration procedure described above. Fit is specifically calculated according to the gaps between predicted and observed data (or "residuals"). Following Shiotsu (2001), who used Rasch analysis to validate the VLT, I will limit the focus here to mean square values only. The two mean square values (infit and outfit mean square) are to be interpreted as to how close they are to the mean of 1. If the value is 1, then there is no discrepancy between the observed and predicted data, so we have exactly the amount of variation that the model predicted. On the other hand, if mean square value is greater than 1, this value shows greater variation than expected; while values of less than 1 show less variation than expected. Both McNamara (1996, p.173) and Bond and Fox (2001, p.178) suggest that a range of 0.75-1.3 is acceptable, with items of mean square values of over 1.3 showing underfit or misfit and those of under .75 showing overfit (i.e., lack of predictability according to the overall pattern of response that the model has generated). However, Bond and Fox also point out other reasonable ranges according to the type of test used; here they widen the range of acceptable values to 0.6-1.4 for rating scale tests and 0.5-

1.7 for clinical observation (Bond and Fox, 2001, p.179). With regard to the present study, it would seem best to follow 0.75-1.3 as suggested by McNamara, on the understanding that we are dealing with rule-of-thumb guidelines rather than absolute cut-off points for deciding misfit and overfit. Some flexibility is warranted then when we come to interpret the mean square values with the collocation production test.

Two other measures provided by the Rasch model are the person reliability index and the item reliability index. The first index is a measure of "the replicability of person ordering we could expect if this sample of persons were given another set of items measuring the same construct" (Bond & Fox, 2001, p.31). In other words, how confident are we that the estimates generated by the model can be reliably inferred? The second index, item reliability, is a measure of how stable the item estimates would stay, were the test to be given to another population. If the item reliability is high, we can be confident that we have a test "in which some items are more difficult and some items are easier, and...we can place confidence in the consistency of these inferences" (Bond & Fox, 2001, p.32).

The final point of familiarisation with Rasch analysis is the type of Rasch model that we should use for analysing the results from the collocation test (hereafter referred to as COLLPROD). As we do not expect the threshold estimates to be the same from one item to the next (i.e., the relationship between person ability and item difficulty will vary from one stimulus item to the next), the partial credit model seems the appropriate choice. This model posits that "each set of threshold estimates is unique to its own individual item" (Bond & Fox, 2001, p.205). At the same time, using the partial credit model obliges us to make one important assumption about the responses that are given by the students. This assumption is that every appropriate response is at a constant

distance from other appropriate responses within the unique set of threshold estimates for each item. The assumption is necessary because there are more than three possible responses for each stimulus item. In the case of **advice**, as we saw earlier, we have a set of over 100 possible appropriate responses, and each response is scored as 1 or 0, according to whether it is within that set of possible appropriate collocates or not. The assumption of constant distance between appropriate responses means simply that any appropriate response is considered as likely as another.

Now that we have exmained the main features of a Rasch analysis for the present study, we will look next at the results from using COLLPROD.

8.4 Results

The main findings of the experiment are reported in this order:

- results from the collocation test
- results from the lexical profiling of collocation responses for the high and low groups
- differences between the high and low group for appropriate collocates
- the relationship between person ability and item difficulty.

8.4.1 Descriptive statistics for collocation appropriacy

The descriptive statistics for collocation appropriacy (i.e., score out of 90) for the whole population (N=89) are shown in Table 8.6 below.

Table 8.6 Descriptive statistics

	Result
N	89
Minimum	17
Maximum	61
Mean	37.93
SD	9.74

The reliability of the test proved to be modest (Cronbach's alpha $\alpha = .78$).

8.4.2 High and low group lexical profiles for all collocates

The collocation responses for the low and high groups were lexically profiled in order to get an overall picture of their collocate production. Table 8.7 below shows the total number of frequent responses. In Table 8.7, one person in the low group had a total of 89 frequent collocate responses, but nobody in the high group had this total. Conversely, four members of the low group produced 84 frequent collocate responses, as did three individuals in the high group.

Table 8.7 Frequent collocate response totals for both groups

Low	Total	High
1	89	
1	88	
1	87	1
4	86	
2	85	
4	84	3
1	83	1
4	82	1
1	81	3
1	80	4
	79	2
	78	1
	77	
	76	
	75	
	74	
	73	2
	72	2

The results showed some overlap between the highest scores of the low group and the lowest scores of the high group. An independent samples t-test showed a significant difference between the two groups for the lexical profile of the students' frequent collocation productions, whether their collocates were appropriate or not ($t = 4.269$, df 38, $p < .001$).

An independent samples t-test confirmed a significant difference between the two groups for the lexical profile of the students' appropriate collocate responses ($t = 5.673$, df 38, $p < .001$), as Table 8.8 below indicates.

Table 8.8 Low and high group appropriate collocate scores

	Low	High
Mean	30.65	46.45
SD	10.20	7.14

8.4.3 High and low group lexical profiles for appropriate collocates only

The two groups' appropriate collocate scores ("hits") were divided into frequent and infrequent items (see Tables 8.9 and 8.10 below), and further t-tests were run on the scores for frequent and infrequent hits.

Table 8.9 Lexical profile of frequent appropriate collocate responses

	Low	High
Mean	29.4	42.35
SD	10.02	6.8

Table 8.10 Lexical profile of infrequent appropriate collocate responses

	Low	High
Mean	1.25	4.1
SD	1.12	2.38

T-test results confirmed a significant difference between the two groups for both appropriate frequent collocation responses ($t = 4.783$, df 38, $p < .001$) and appropriate infrequent collocation responses ($t = 4.844$, df 38, $p < .001$).

8.4.4 Differences between high and low group collocate responses

To explore more finely the differences between the two groups, two further analyses were carried out. First, all appropriate collocates were sub-divided by word class into

adjectival, *nominal*, *verbal*, *nominal-verbal* (i.e., homonyms such as **answer** and **play**), and *other* (mainly prepositions and adverbials). A t-test was then run for each of these word classes. A second similar analysis was completed for frequent appropriate collocates only (as the low group did not produce enough infrequent appropriate collocates to warrant a further comparison). The results are reported in Tables 8.11 – 8.13 below. Table 8.11 shows the mean scores for all appropriate collocates, Table 8.12 for frequent appropriate collocates, and Table 8.13 reports differences for specific word classes.

Table 8.11 High and low group means: all appropriate collocates

Word class	Low	High
adjectival	14.95	16.50
nominal	5.95	9.60
verbal	6.00	13.15
nominal-verbal	2.25	4.75
other	1.55	2.35

Table 8.12 High and low group means: frequent appropriate collocates

Word class	Low	High
adjectival	14.15	14.80
nominal	5.60	7.60
verbal	5.95	12.45
nominal-verbal	2.25	4.50
other	1.55	2.35

Table 8.13 Differences between high and low means for word class choice

ALL appropriate collocates	
verbal	$p < .003$
nominal-verbal	$p < .001$
FREQUENT appropriate collocates	
nominal	$p < .003$
verbal	$p < .002$
nominal-verbal	$p < .001$

Overall, the two groups showed significant differences in frequent nominal, verbal, and nominal-verbal collocates.

8.4.5 The relationship between person ability and item difficulty

The person reliability index for COLLPROD was .77 and the item reliability index was .93. The first result is very similar to Cronbach's alpha (.78), but the second result is much more robust. We have an initial indication that the instrument has a good spread of items from difficult to easy, but we can be less confident that the same learners would respond with the same consistency, were they to be tested on a similar measure with different items. Results for item fit should therefore be very good, while we can expect the findings for person ability to show some degree of variation.

As mentioned above in the explanation of the Rasch model, we will limit our focus on item difficulty to the infit and outfit mean square values. Here we are interested in two points. First, we want to see whether the infit and outfit mean square values are located within the rule-of-thumb range of 0.75-1.3. If they are not, then we can identify these items as problematic. Second, we wish to see whether any items have a mean square value of 1, i.e., whether they have perfect fit. We would not want too many items in the test to have such a zero range of variability. The results for item fit are presented in Table 8.14 below.

Table 8.14 Item fit results for COLLPROD

Item #	Word	Infit Mean Square	Outfit Mean Square	Infit *t*	Outfit *t*
6	**DECISION**	0.85	0.85	-1.25	-1.18
12	**HEALTH**	0.87	0.85	-1.11	-1.21
19	**POLICE**	0.88	0.87	-0.98	-1.01
23	**REASON**	0.88	0.88	-0.91	-0.95
5	**DEATH**	0.9	0.86	-0.68	-0.94
25	**ROLE**	0.91	0.9	-0.67	-0.74
10	**FRIEND**	0.93	0.94	-0.48	-0.46
7	**EXAMPLE**	0.94	0.93	-0.47	-0.52
26	**SUPPORT**	0.95	0.97	-0.28	-0.18
8	**EXPERIENCE**	0.96	0.93	-0.29	-0.48
14	**INTEREST**	0.96	0.95	-0.22	-0.3
20	**POWER**	0.96	0.99	-0.27	-0.03
24	**RESEARCH**	0.96	0.9	-0.15	-0.5
9	**FAMILY**	0.99	1.01	-0.01	0.14
21	**PROBLEM**	0.99	1.01	-0.04	0.1
30	**WORK**	1.0	0.97	0.08	-0.13
1	**BODY**	1.0	1.0	0.03	0.06
22	**QUESTION**	1.0	1.0	0.08	0.03
4	**COUNTRY**	1.02	1.0	0.22	0.04
3	**CHILD**	1.02	1.01	0.2	0.1
2	**CAR**	1.03	1.03	0.29	0.28
13	**HOUSE**	1.03	1.04	0.28	0.37
27	**VALUE**	1.07	1.06	0.56	0.51
15	**ISSUE**	1.08	1.15	0.62	1.1
17	**LIFE**	1.09	1.09	0.76	0.75
11	**GOVERNMENT**	1.1	1.15	0.85	1.16
28	**VOICE**	1.13	1.13	0.98	1.04
16	**LAW**	1.13	1.15	1.04	1.16
18	**PAPER**	1.14	1.2	0.97	1.34
29	**WAR**	1.23	1.29	1.73	2.1
	Mean	**1.0**	**1.0**	**.03**	**.06**
	SD	**.09**	**.11**	**.71**	**.82**

Table 8.14 shows that all items have acceptable mean square infit and outfit values, and that three items (**body, question** and **work**) have perfect fit. At this point in the analysis, we do not know the reason for such zero variation. That will become clearer when we consider the item-ability map for COLLPROD (see Figure 8.1 further below), but first I will present the results for person ability, as shown in Table 8.15 below.

Table 8.15 Person fit results for COLLPROD

ID	Infit Mean Square	Outfit Mean Square	Infit *t*	Outfit *t*
9	0.5	0.52	-2.77	-2.55
34	0.51	0.51	-2.67	-2.6
35	0.55	0.54	-2.17	-2.22
77	0.58	0.57	-2.17	-2.2
11	0.6	0.62	-1.81	-1.63
46	0.64	0.63	-1.79	-1.81
89	0.64	0.64	-1.83	-1.76
6	0.68	0.7	-1.57	-1.44
14	0.69	0.72	-1.5	-1.31
4	0.71	0.77	-1.28	-0.93
1	0.72	0.71	-1.32	-1.34
57	0.76	0.7	-0.87	-1.1
37	0.76	0.77	-1.09	-1.0
8	0.76	0.79	-1.05	-0.86
2	0.77	0.79	-1.05	-0.93
10	0.79	0.77	-0.93	-1.03
25	0.8	0.8	-0.82	-0.8
70	0.81	0.78	-0.86	-0.98
80	0.81	0.8	-0.87	-0.88
16	0.81	0.81	-0.82	-0.85
27	0.81	0.83	-0.86	-0.74
12	0.81	0.84	-0.84	-0.7
65	0.82	0.77	-0.79	-1.02
36	0.83	0.8	-0.7	-0.82
20	0.83	0.86	-0.75	-0.6
30	0.83	0.87	-0.73	-0.53
53	0.84	0.85	-0.69	-0.62
79	0.85	0.86	-0.67	-0.57
85	0.87	0.81	-0.56	-0.84
49	0.87	0.86	-0.52	-0.54
48	0.88	0.89	-0.43	-0.36
66	0.88	0.89	-0.5	-0.42
58	0.89	0.87	-0.45	-0.52
44	0.91	0.88	-0.36	-0.48
67	0.91	0.89	-0.25	-0.31
68	0.91	0.91	-0.35	-0.32
24	0.93	0.9	-0.25	-0.37
72	0.93	1.05	-0.23	0.29
33	0.94	0.86	-0.16	-0.48
87	0.94	0.99	-0.21	0.04
17	0.95	0.95	-0.17	-0.17
21	0.96	0.93	-0.12	-0.22
81	0.97	1.02	-0.08	0.18
47	0.98	0.94	0	-0.18
82	0.98	0.94	-0.04	-0.2

29	0.98	1.01	-0.01	0.12
40	0.99	0.97	0.05	-0.07
42	0.99	0.99	0.01	0.03
45	0.99	1.01	0.05	0.12
54	1.00	1.07	0.05	0.38
43	1.02	0.99	0.16	0.05
32	1.02	1.03	0.18	0.21
78	1.02	1.04	0.17	0.24
62	1.04	1.04	0.27	0.26
41	1.05	1.04	0.31	0.27
83	1.05	1.09	0.3	0.44
26	1.07	1.07	0.38	0.38
13	1.09	1.07	0.46	0.37
7	1.1	1.1	0.53	0.49
76	1.1	1.11	0.5	0.54
50	1.11	1.05	0.55	0.31
28	1.11	1.13	0.57	0.63
51	1.12	1.04	0.57	0.23
88	1.12	1.12	0.62	0.6
56	1.14	1.1	0.65	0.5
15	1.15	1.16	0.71	0.74
86	1.16	1.1	0.77	0.5
63	1.2	1.23	0.83	0.91
31	1.21	1.19	0.96	0.86
18	1.21	1.23	0.89	0.95
84	1.22	1.23	1.0	1.02
52	1.23	1.12	0.87	0.48
71	1.23	1.23	1.05	1.03
3	1.23	1.26	1.02	1.14
22	1.26	1.24	1.13	1.05
69	1.28	1.35	0.98	1.13
23	1.29	1.31	1.28	1.35
61	1.31	1.19	1.24	0.8
75	1.32	1.27	1.34	1.14
64	1.32	1.38	1.14	1.28
59	1.33	1.27	1.38	1.15
55	1.34	1.38	1.39	1.48
19	1.37	1.5	1.54	1.96
5	1.45	1.44	1.79	1.73
73	1.48	1.61	1.96	2.36
38	1.53	1.48	2.15	1.94
60	1.57	1.51	2.11	1.85
39	1.6	1.54	2.39	2.16
74	1.82	1.82	3.08	3.0
Mean	**1.0**	**-.01**	**1.0**	**-.01**
SD	**.26**	**1.1**	**.26**	**1.1**

If we adjust the rule-of-thumb range of acceptability for mean square values to 0.65-1.4,

person misfit involves seven learners. I have extracted the person misfit group in Table 8.16 below.

Table 8.16 Person misfit results for COLLPROD (infit mean square < .65)

ID	Infit Mean Square	Outfit Mean Square	Infit t	Outfit t
9	0.5	0.52	-2.77	-2.55
34	0.51	0.51	-2.67	-2.6
35	0.55	0.54	-2.17	-2.22
77	0.58	0.57	-2.17	-2.2
11	0.6	0.62	-1.81	-1.63
46	0.64	0.63	-1.79	-1.81
89	0.64	0.64	-1.83	-1.76

Person misfit involves seven learners, with the direction of "the misfit going from instrument to the person" (McNamara, 1996, p.178). This comes to 8% of the population, which is much higher than the 2% benchmark which McNamara gives as the upper limit for the level of person misfit. In such cases, McNamara argues that the instrument will need to be revised in order to reduce the level of person misfit—a point we will return to in the discussion section.

The final part of the Rasch analysis results concerns mapping person ability and item difficulty. Figure 8.1 below shows this map for COLLPROD. On the left of the figure is the logit scale, which goes from -2 to +2 (a 4-unit range). Average item difficulty has been fixed at 0. The top of the scale indicates increased item difficulty and stronger collocation ability. Conversely, the bottom of the scale points to greater item facility and weaker collocation ability. Each X represents two learners.

Logits	Persons	Items
2.0		
		research
1.0		
	X	
	X	body
	X	support
		country death interest power
	XXX	
		role
	XXX	value war work
	XXXXXXXX	government police
	X	family
0	**XXXXXXX**	experience issue
	XXXXXXXX	child health law paper
	XXXXXX	car decision home
	XXXXXXXXXXXX	reason
	XXXXXX	example
	XXXXXXX	
	XXXX	life
	XX	voice
	XXXXX	friend problem
	XXX	
-1.0	**X**	
	XXXX	
	X	question
	XXX	
	X	
	X	
-2.0		

Figure 8.1 Item-ability map for COLLPROD

Figure 8.1 shows that **research** turns out to be the most difficult item to collocate, and **question** the easiest. The two items, **experience** and **issue**, prove to be of average difficulty to collocate. A person of 0 logits ability has a 50% chance of being to

collocate **experience** and **issue**. The logit scale lets us see that, with items at about 1 logit more than a person's ability (here, **body** and **research**), the chance of being able to collocate these words appropriately is around 25%, whereas, with items at about 1 logit less than a person's ability (that is, **friend**, **problem** and **question**), the chance of being able to collocate these words appropriately increases to around 75%.

Although the spread in the top half of the map is not quite an exact mirror image of the bottom half, Figure 8.1 shows that the spread of item difficulty is even: There are 14 items above 0 logits, and 14 items below. This explains why the item reliability index was so strong at .93. On the other hand, Figure 8.1 also tends to indicate person (collocation) ability is unevenly spread (i.e., there are more learners responding in the bottom half of the figure than the top)—hence, the more modest result for the person ability index at .77.

8.4.6 Summing up the results

To summarize the main findings, we have been able to identify several significant differences in collocation production between low-intermediate and advanced students of English. We found that these differences held good along a number of dimensions, namely (a) the overall score for appropriate collocates, (b) the infrequent collocate score, (c) the frequent collocates score, and (d) particular word classes. Through Rasch analysis, we have also managed to determine that COLLPROD has a surprisingly strong item reliability index, and that there is a marked degree of variation in person ability in supplying appropriate collocates.

8.5 Discussion

The basic goal of this study was to develop an expanded instrument suitable for testing

L2 collocation knowledge. We restricted the items to highly frequent nouns and piloted these on both NS and NNS before the main test administration. By eliciting a maximum of 90 collocates per subject, the modified measure provided large individual data sets for analysis. Both the lexical profiling and the word class analysis of the data show that more collocationally competent learners produce significantly more infrequent *and* frequent collocates. Not only are their L2 lexicons bigger, but core items such as the 30 nouns used in the test also appear to be more readily linked with other highly frequent lexical items, as well as with infrequent vocabulary. The second insight that this test offers is that adjectival links form the dominant type of collocation for less proficient learners. These responses account for 50% of the responses of the lower group, but only a third of the responses of the higher group. Furthermore, we seem to have some evidence that, in later stages of collocation development, adjectival links remain frequent, but are also complemented by nominal and verbal, as well as nominal-verbal links. In other words, learners still produce adjectival collocation links, but appear to diversify these with other types of word class connections in the collocations that they produce.

We have grounds then for claiming that this modified test of L2 collocation production is an improvement. The elicitation of large individual data sets is a particular strength of COLLPROD. If we recall that one of the most commonly cited studies of collocation production (i.e., Bahns & Eldaw, 1993) worked from just 15 V + N collocations, we have come quite a way further in this chapter in addressing issues of validity and reliability. In terms of validity, we have used a large number of items to elicit data and to allow for different appropriate responses to the same items. We have not limited the subject response to one correct answer, but rather explored how subjects'

L2 collocation knowledge can be extensively and flexibly sampled. In terms of classical reliability, COLLPROD performs modestly well with Cronbach's alpha at .78. Although, at first sight, it would seem that the reliability of COLLPROD could be further improved if poorly performing items were replaced, the Rasch analysis suggested otherwise. The item reliability index is very high .93, but the person reliability index is again modest at .77. Rasch analysis also showed that only three items had perfect fit (**body, question** and **work**), and there is no reason to believe that such items need to be replaced. It is, however, possible that Cronbach's alpha and the person reliability index would improve, were COLLPROD to be run with a larger population. McNamara, for example, notes that recommended sample sizes should comprise 100 subjects minimum, although he adds that smaller groups can be used 'if the size of the error term is not an imperative consideration' (p.163). In the case of the present study, a population of 87 would seem to be just about the lower limit for this 30-item test.

Two other improvements are worth comment. The first is that COLLPROD has allowed us to pin down more precisely some of the individual differences that we noted in Chapter 7. As a result, we have established several patterns of significant variation between groups of individuals of high and low proficiency. The second is that both the lexical profiling and word class categorization of appropriate collocate responses have let us move beyond a strictly lexical analysis of L2 collocation knowledge and see collocation links in macro-grammatical terms too. This is a level of delicacy that Zhang (1993) and Gitsaki (1996) pursued through parsing their essay data for huge numbers of pre-established lexical and grammatical types of collocation (66 and 37, respectively). The present study has empirically arrived at some significant results for basic word

class choice rather than pre-established collocation type.

L2 collocation knowledge is generally thought to be an advanced type of lexical knowledge, but the results from this study suggest otherwise. COLLPROD shows that the development of L2 collocation knowledge can be tracked at both pre-intermediate and advanced levels of proficiency. Part of the reason for this is that COLLPROD uses individual nouns to elicit L2 collocation knowledge, but, unlike previous studies of L2 collocation knowledge (cf., Schmitt, 1999), does not analyse the results in terms of "depth of knowledge" of the particular stimulus word. Rather, both the lexical profiling and word class analysis of appropriate collocate responses provide representative samplings that shed light on a more general understanding of the L2 lexicon. In their original work on lexically profiling word associations, Meara and Fitzpatrick (2000) claimed that their test, Lex30, provided a "practical index of productive vocabulary." With COLLPROD, we have some reason to believe that a "practical index of L2 collocation production knowledge" has been achieved: COLLPROD relates L2 collocation proficiency to general English ability within the two quite clearly distinguished bands of pre-intermediate and advanced. In more precise terms, COLLPROD lets us see that L2 collocation development is a function of *ongoing L2 lexical re-organisation, not just growth in terms of size*, through the establishment of:

- greater links between frequent lexis
- greater links between frequent and infrequent lexis
- word-class diversification/an increase in polysemous links.

These insights are important for developing our understanding of the L2 lexicon.

There are, however several limitations with COLLPROD that should be mentioned. The test is partially specific to its context of use in that an item such as **law**

may well be more familiar to Faculty of Law students than other populations. Both the low and high groups could produce collocates such as **criminal, constitutional, faculty** and **international**. A future version of COLLPROD might need to replace such domain-specific items with more general ones. A further weakness is that collocation databanks become quickly out of date and may not pick up on newly coined collocations in mass circulation and re-use in print and virtual media. A case in point here is the fact that **terror** does not occur as a collocate of **war**. Ironically, this means that it is quite possible for learners who regularly read the news in English to score lower on certain items than those who don't. Another concern is that, while data sets of 90 collocates can be considered a major improvement on previous tests of productive L2 collocation knowledge, they nevertheless constitute a relatively limited sampling of the L2 lexicon. Thus, the claims that we have made about the usefulness of COLLPROD need to be treated with some degree of caution. An important future stage in the validation of this test may involve exploring learners' test-taking processes through think-aloud interviews; this would also help us to start identifying possible relationships between responses on this experimental measure and learners' real-world productive collocation knowledge. A final limitation follows from the low mean score of 46.45 for the high group. Seen in negative terms, this result may suggest that the test is rather too difficult, but it is not quite clear whether we should make such an assumption, given the other evidence that the Rasch analysis provided. Again, further validation of COLLPROD in the future would be helpful here.

8.6 Conclusion

In this chapter, we have adapted and expanded a stimulus-response measure of L2 collocation knowledge. The results from the study show that COLLPROD is a robust

improvement on the experimental work reported in Chapter 7 where we initially explored how a scaled-down version of such a measure might work. The present instrument has enabled us to distinguish the L2 collocation knowledge of low and high proficiency learners and identify several interesting differences in the way that the L2 mental lexicon is organised at those two stages of development. Yet, COLLPROD does not let us understand how learners address the challenge of improving their collocation knowledge in practice. We have extensively analysed the results from this test, but we have viewed at a distance what learners themselves do to develop their L2 collocation ability. We need, I believe, to re-consider our options and perhaps accept at this point that we have done what we can to understand L2 collocation knowledge in quantitative terms. At some point, it lies in our interest to engage in a well-planned longitudinal qualitative study so that we can develop a real sense of what learners do or don't do, and what problems they face, in becoming more collocationally competent in a second language. In the final experimental chapter, we will therefore explore qualitative characteristics that may underpin the development of L2 collocation knowledge.

Chapter 9

Exploring Individual Development of Collocation Knowledge

9.0 Overview

Most of the work presented so far in this thesis has featured large-scale interventions. In seven of the previous experiments, we have developed our understanding of collocation recognition and production through considering results from research with large numbers of subjects. Although in one experiment we took a different tack and considered a longitudinal study with three students over one academic year, the focus has nevertheless been consistently quantitative. With these previous experiments, our broad concern has been to establish what learners know in linguistic terms. In the last experiment, we found significant differences in collocation production between a low proficiency group and a high proficiency group, and we noted how more collocationally proficient learners not only produce more collocations, but also differ in the type of word combinations that they make. In contrast, the investigation in this chapter is concerned with how learners come to know what they know and to use what they know. In this study, I explore in detail how individual learners organise, understand and interpret their own development of lexical and collocation knowledge over time. This exploration considers not just linguistic, but also cognitive and meta-cognitive dimensions of second language collocation ability, situated within the context of a content-based English course. In this chapter, I will focus on the same three students whose collocation development was quantitatively examined in Chapter 7, but I will draw on extensive interviews conducted with Emi, Kensuke and Shoichi over the same academic year.

9.1 Introduction

By providing snapshots of lexical knowledge at single points in time, vocabulary tests seek to establish quantifiable and significant differences in lexical knowledge between populations which are often characterized by broad polar markers such as low/high proficiency, EFL /ESL and non-native/native. While the results may allow for generalization of similarities and differences between representative groups, particular patterns of individual vocabulary acquisition and development necessarily remain somewhat obscured. Research into vocabulary learning strategies has attempted to redress the balance by exploring how students learn vocabulary in a second language. Yet, it has sometimes suffered from the same drawback of providing large-scale quantifiable results that are difficult to relate to learners' individual experiences of learning vocabulary *in situ.* A study such as Schmitt (1997) illustrates this tension between the quantifiable and the qualitative. Schmitt surveyed 600 Japanese learners about their use of 40 different English vocabulary learning strategies and examined the results in terms of the polar categories *most used / least used* and *most helpful / least helpful.* The study broadly confirmed that patterns of strategy use changed over time as learners moved through the education system—put simply, that learners tended to shift from a preference for form-focused to meaning-focused strategies. However, Schmitt notes that his study "does not shed any light on *why* the patterns of usage change" (p.224).

This gap suggests three important weaknesses with such a large-scale taxonomic survey. The first problem is that it does not let us understand whether and how learners combine different strategies. So, we have no sense of the interplay between discrete components of individual learning process and development. A second related problem is

that Schmitt's study does not account for context of learning except with the broad categorisation by age of learners as junior high school, high school, university, or adult. Although the results show that, across all four age categories, *using semantic maps* is among the least used strategies and *taking notes* among the most used, it is quite conceivable that these two strategies would not be so divergently positioned in specific learning contexts. Suppose, for example, that students are expected to research and *make* notes for themselves rather than simply *take* notes of what a teacher has explained. In such a situation, we might well find that two strategies previously presented as unconnected became more strategically exploited and linked as part of an overall research process and engagement with content through English. A final weakness in Schmitt's paper is that the investigation does not examine how specific strategy use helps particular learners succeed or fail in reaching their learning goals and aims. *Written repetition* may, for instance, be the strategy most used by junior high and high school Japanese learners, but does it actually help learners successfully develop their vocabulary, not to mention collocation ability?

Other studies have drawn out contrasts between what vocabulary strategies successful and unsuccessful learners use and how they combine them. These studies have featured both small-scale qualitative and large-scale quantitative inquiries, and, in a series of investigations, Gu has in particular tried to reconcile the tension between the quantifiable and the qualitative. Gu (1994) examined how two learners—"one 'good', the other 'poor'" (p.378)—dealt with intensive reading and vocabulary learning. The poor learner saw learning vocabulary in terms of remembering form and meaning, whereas the good learner was concerned with putting words into context, using them and seeing what collocational patterns particular words had. The poor learner was mostly concerned with memorizing

new words, while the good learner was much more meta-cognitively active and evaluative in considering which particular strategies would be most useful for a particular purpose. In a later large-scale experiment with 850 Chinese university students, Gu and Johnson (1996) were able to develop different profiles of successful and unsuccessful vocabulary learners by correlating results from an English vocabulary learning questionnaire with English proficiency. They found that the two most successful groups of learners were *readers* and *active strategy users*. *Readers* favoured developing their vocabulary through reading and careful study rather than by memorization, whereas *active strategy users* spent more time and effort learning English vocabulary and were highly varied in the strategies that they used:

> They guessed more, used more dictionary strategies to learn vocabulary, took more notes, did more memorization, and activated more newly learned words than their peers. (Gu & Johnson, 1996, p.664)

The least successful group, *passive strategy users*, had limited ideas about how to learn English, showed weak motivation, and made little self-initiated effort to develop their vocabulary.

Gu and Johnson's main findings match those of Kojić-Sabo and Lightbown (1999) who explored the vocabulary learning strategies of two groups of learners in different contexts—one ESL, the other EFL. They found that success in both cases was underlined by *learner initiative* and *independence*, as well as by *time* spent outside the classroom on vocabulary learning. Both groups were found to view *note-taking* and *dictionary use* as important, but the EFL group placed a higher priority on *review* than the ESL group. The review techniques favoured by the EFL group included self-oriented activities such as reading and re-reading notes and quizzing themselves about new vocabulary; they also

involved other-oriented tasks such as "cooperative reviewing activity with their friends" (p.183). On the other hand, ESL learners are described as being more creative in their review techniques, partly because of their context of learning. They could, for example, review new vocabulary items "by using them in their daily conversations" (p.183). Kojić-Sabo and Lightbown are careful to emphasize that not all strategies are "universally valid or useful to all learners" (p. 190) and that learners will develop their own individualised vocabulary learning preferences influenced by personality, motivation, and other factors. One final point from the study is that the importance of meta-cognitive awareness and active involvement in vocabulary learning matches insights from more general studies of self-directed autonomous language learning (e.g., Benson, 2001; Dam, 1991; Little, 1991; Wenden, 1991) and cognitive theories of language learning that account for individual difference (e.g., Skehan, 1989, 1998).

The critical role of meta-cognitive awareness and active involvement is further explored by Gu (2003) in his analysis of the vocabulary learning strategies of two successful Chinese learners of English, both of whom were non-English major first-year university students taking an intensive reading English course. Following Gu and Johnson (1996), Gu identified one of the learners as a *reader* and the other as an *active strategy user*. Despite this specific characterization of difference, Gu also noted several meta-cognitive and cognitive similarities. At a meta-cognitive level, the two successful learners both showed self-initiation in making the effort to develop their English vocabulary. They also used different strategies at different times with different words, chose vocabulary according to conscious criteria such as interest and relevance to text comprehension, and sought out multi-word expressions. At a cognitive level, the two learners employed an array of

strategies, including word list memorization and repetition; they also exploited dictionaries both to understand words and to learn them by making notes. They were furthermore keen to use English, as well as the vocabulary that they had chosen to learn. Gu concludes:

> ...what makes vocabulary learning an art for these students is not the strategy repertoires they used or how often they used them but the flexible and skilful analysis, choice, deployment, execution, and orchestration of all strategies at their disposal in accordance with their own preferred style of learning. (Gu, 2003, p.99)

His conclusion echoes the development of individualised vocabulary learning preferences that Kojić-Sabo and Lightbown also highlighted.

This review of the literature on vocabulary learning strategies and concomitant factors reveals some consensus about how successful and unsuccessful learners in different contexts organise their vocabulary learning. These studies have involved both small-scale qualitative and large-scale quantitative investigations, with Gu's work in particular focusing on balancing the generalizable with the particular and localized. That said, these strategy-focused investigations have had very little to say about the development of second language collocation knowledge. However, this is not necessarily a problem if we assume that the development of collocation ability is part of the larger construct of lexical learning. Such an assumption would lead us to expect to find broadly similar patterns of process with successful and unsuccessful collocation learners. Our expectations would at least include:

(a) more successful collocation learning (i.e., development) involves a high degree of meta-cognitive awareness and control of both collocation recognition and production;
(b) more successful collocation learning involves explicit effort and motivation, particularly with regard to reading extensively and/or employing a variable set of strategies selected according to individual goals and plans in the interlinked processes of learning, reflection and communication;
(c) more successful collocation learning involves time outside classroom

learning, note-taking, dictionary use, and review, where review tends to be both self-oriented (i.e., private and individual/independent) and other-oriented (i.e., social and interdependent);

(d) less successful collocation learning involves a predominant focus on form and accuracy that downplays a concern with fluency and meta-cognitive learning decisions;

(e) less successful collocation learning involves a predominant focus on memorization of L2 forms to the exclusion of a concern with meaning and use.

These five expectations are presented here as a minimal set of points of reference for the present study.

9.2 Research focus

In Chapter 7, I reported on a quantitative analysis of the three students' word association and collocate production. Overall, the results from that study showed that (a) Emi made clear and consistent progress; (b) Kensuke made little progress; and (c) Shoichi made fluctuating but positive progress. In this parallel study, I am interested in moving beyond these broad generalisations and making a much more detailed and contextualised interpretation of their English vocabulary and collocation development. Using the minimal set of expectations presented above, I take a grounded theory analysis approach to the interview protocols to describe and explain similarities and differences between the three learners. Where possible, then, theoretical insights may emerge from the data rather than be imposed a priori on the inquiry.

9.3 Method

9.3.1 Participants

The three students who were the subjects for the longitudinal experiment reported in

Chapter 7 were also interviewed about how they managed learning collocations. Emi, Kensuke and Shoichi's previous experiences of learning English had consisted mainly of reading and translation classes, though they each had taken English communication and discussion courses earlier at university. Both Shoichi and Kensuke had also been learning Chinese as their second foreign language, whereas Emi had been taking courses in German and French. They had no prior experience of learning English collocations as an intrinsic part of their vocabulary development. Although the interviews took place over the course of one academic year, the first semester was used as a period of familiarization in which an appropriate interview format could be developed. Thus, only the interviews from the second semester are considered here.

During the second semester, three interviews were held with each student (November, December and January). The interviews took place at the end of each 1-month research-discussion-presentation cycle of the course. At the start of a cycle, the students chose their own human rights issue to work on. They researched, found English sources and made notes outside class, reported and explained their research in class, and gave a 10-minute poster-presentation in the final lesson of each cycle. The students also handed in a portfolio of their work at the end of a cycle. The portfolios included copies of the sources that had been used, the students' notes on those sources, and notes from class discussions and preparation of individual presentations. The portfolio materials were used to provide specific documented examples of the students' learning. On average, each interview lasted between 40 and 60 minutes.

9.3.2 Instrument: organisation of interviews

The interviews followed a semi-structured format and were organized around the following broad themes:

- What did you learn in the previous cycle?
- How did you organise your vocabulary learning?
- What problems or difficulties did you notice in how you organized your vocabulary learning?
- What changes do you notice in your vocabulary learning in this cycle?
- How do you evaluate those changes?
- Any other comments?

These starting questions were designed to enable the students to (a) describe what they had researched so that they could situate their vocabulary learning within a broader frame of self-directed content-based learning in English; (b) identify problems that they had faced in their English vocabulary learning, with a particular focus on collocations; (c) discuss changes in their vocabulary and collocation learning as part of their overall development over the academic year; (d) evaluate those changes critically; (e) have an opportunity to raise other questions and issues if they wanted to.

The semi-structured format allowed for clear and consistent stages to be established in which I could respond with specific follow-up questions to what each student said.

9.3.3 Overview of grounded theory analysis

Analysis followed a grounded theory approach. This can be succinctly summarized in these terms:

> In overall terms the analyst typically works from an initial topic or research question(s), to data gathering, through initial treatment of unstructured materials...possibly more data gathering and analysis, and to a set of theoretical categories, interpretations, models, and written accounts of theory...this flow is accompanied by a gradual development of conceptual focus away from local descriptions inherent to the data toward more ordered

> and analytical (i.e., theoretical) concepts and categories. (Henwood & Pidgeon, 2003, p.136)

Three main phases of analysis can be identified for dealing with grounded data: description, conceptual ordering, and theorizing (Strauss & Corbin, 1998). Describing refers to representing in detail the data "without stepping in" (Strauss & Corbin, 1998, p.25). In the case of interview materials, description would equal basic transcription. Conceptual ordering involves classifying phenomena in the data into explicit categories, while theorizing is based on identifying relationships between categories and developing a systematic explanation of such connections. With interview transcripts, then, the second stage would involve analyzing turns and sequences, labeling these broadly, and then moving towards establishing more explicit category definitions. The third stage of theorization would focus on examining interconnections between categories, so that a grounded theoretical interpretation can be made.

Strauss and Corbin also present a number of techniques for examining concepts, grouping them into categories, and relating categories to each other. These include systematic comparison of categories and explicitly defining categories. An important aspect of these analytical processes is to provide ways for the researcher to move beyond 'common sense' classification and theorization, which Silverman (2000; 2003) identifies as a potential threat to grounded theory analysis. So, although grounded theory analysis focuses on the effort to develop new theoretical insights from initially unstructured data, there is clearly a need to keep a background set of reference points to help inform and define categories.

To address these issues in practice, I will, in the rest of this section, describe in some

detail the particular stages in the analysis and explain important decisions that I made as I worked through the data.

9.3.4 Stages of analysis

All the interviews were audio-recorded and transcribed (see Appendix 9.1 on the CD-ROM for complete transcripts). Initially, single participant turns or sequences of turns were coded with a broad conceptual category such as *Research process*, *Vocabulary*, *Dictionary use*, and *Collocation.* Later, each exponent of a broad category was then amplified with particular sub-labels. For example, if one of the students explained what they had learned about the human rights issue that they were investigating, the initial broad category of *Research Process* was recoded as *Research process: Understanding of the issue*; where a student gave an example of a particular word that they knew and commented that they didn't know its collocates, the initial category of *Collocation* was re-labelled as *Collocation: Word known but collocates unknown.* As the instances of such more specifically defined categories multiplied, I then started to add *Example 1*, *Example 2*, and so on at the start of the category label (i.e., *CollocationExample1: Known word but collocates unknown*), so that I could then sort and re-order the data sequentially by category labels to observe similarities and differences between exponents of the same conceptual category.

The steps in the analysis that I have just outlined allowed me to deal with description and conceptual ordering in a consistent manner. However, it is important to note that the re-ordering of the data broke the original linear sequence in the development of individual interviews; as such, it seemed to risk obscuring other salient elements of each individual's

representation of what they had been doing. Although Silverman is quite insistent that a conversational analysis should not try to "make sense of single lines of transcript or utterance in isolation from the surrounding talk" (2000, p.151; 2003, p.358), it is difficult to see how a grounded analysis of interview data can be conducted without exploiting both linearity and non-linearity in the conceptual ordering stage. I decided to deal with this tension by re-reading each interview transcript to see whether the categories that I had created non-linearly matched the flow of what each interviewee said at particular stages. This helped me question again the categories that I had been working with and move beyond assumptions that I had formed during the conceptual ordering. Where I decided that I had missed something of importance, I adjusted the previous coding and sorted the data again. For example, as I compared again the emerging categories with the transcripts, I noticed that "care" was an appropriate umbrella term for how the three students talked about giving attention or not to what they were doing for themselves and for others. I had earlier picked up on "explain to self & others" as a conceptual category, but as the data became more structured, I made the initial category a sub-component of the new one. In such cases, several important adjustments were made to the labelling of exponents and categories already completed.

The revision led to 10 core categories initially. These were:

(a) care
(b) dictionary use
(c) engagement
(d) notes
(e) noticing
(f) problematizing learning
(g) reading
(h) sources of information
(i) vocabulary connections
(j) vocabulary selection.

The analysis showed that two of these categories, *sources of information* and *notes*, were sparsely developed and did not seem worth, for the time being, developing further. I had

evidence in the portfolio materials of what sources of information the students had used, and what notes they had made, so, for the sake of simplicity, I decided to exclude these categories from further analysis. (I will however refer to these two areas briefly in the later discussion section of this chapter.)

The eight remaining core categories could then be defined, as shown in Table 9.1 below. I have sorted them into three groups, contextual, lexical, and meta-cognitive.

Table 9.1 Grouping and definition of core categories

Group	Category	Definition
1 Contextual	**A Care**	• *Concern for and interest in (or lack of) learning for self and communicating in English with others*
	B Engagement	• *Involvement in trying to understand and to develop further understanding of a particular human rights issue through English*
2 Lexical	**A Reading**	• *Using texts to find information and identify important ideas and vocabulary/collocations*
	B Vocabulary selection	• *Choosing vocabulary from text for further action*
	C Vocabulary connections	• *Using different strategies to connect selected vocabulary to own knowledge and planned collocation use*
	D Dictionary use	• *Using dictionaries (English-Japanese, Japanese-English, monolingual English, monolingual English collocation) to take further actions with vocabulary/collocations*
3 Meta-cognitive	**A Noticing**	• *Perceiving vocabulary/collocations in text and making informed judgements about what further actions to take*
	B Problematization of learning	• *Developing a sense of past English learning preferences, questioning them, and developing new ways of learning English according to individual experience, awareness and goals*

The first group, contextual, places together *care* and *engagement*. These two categories suggest that communication and involvement have an important influence on the

development of L2 collocation knowledge. Although elements of both categories are particular to the local context, they may also potentially point to some universal factors important in an EFL environment where learners have restricted exposure to English. The second group of categories consists of *reading*, *vocabulary selection*, *vocabulary connections*, and *dictionary use*. I have classified these as lexical, because it is within these categories that the bulk of lexical decision-making and actions that may foster or hinder collocation development take place. The third group, meta-cognitive awareness, includes the two categories of *noticing* and *problematization of learning*. The meta-cognitive dimension involves not just learners' actions in the present (i.e., noticing), but also their changing sense of their development and English learning histories. Without transforming their understanding of learning and using English, it is uncertain whether collocation development will be initiated and sustained in an EFL environment where classroom contact hours are severely limited.

In the presentation of results that follows, I have decided to maximize the reporting of the second group of categories, namely *lexical* as it is the lexical domain that is of central relevance to the scope of this study.

9.4 Results

The main findings from the analysis are reported by category for each individual, in this order: Kensuke, Shoichi, Emi. In most cases, this allows for an increasingly complex presentation of each core category to be made. The core category definition is re-introduced at the start of each sub-section, and the individual constructs of each category are either (a) visually displayed with a brief commentary for each learner or (b) explained

in text only, with some visual displays to be found in Appendix 9.2: Visual displays of certain other key categories (cd-rom). A common theorization of collocation development across individuals is presented in the final part of this section.

9.4.1 Care

This core category is defined as "concern for and interest in (or lack of) learning for self and communicating in English with others." In a course that is based on individual research projects over 4-week cycles, concern for learning and communication becomes constructed by learners as part of the general context of situation within which they may develop their L2 lexical and collocation knowledge. The following learner constructs show how the more the students become concerned with their own learning and communicating what they learn to the others in class, the more they are likely to be able to self-initiate and sustain vocabulary/collocation development. The core category has two aspects: self and others. While learning may be seen as individual and independent, development of learning is social and interdependent and achieved through communication with others.

Kensuke constructs care in the following manner:

care _____ others _____ present _____ be accurate

Figure 9.1 Care for overall process in Kensuke's account

His construct of care is minimally developed. In presenting to others what he has learned, Kensuke is solely concerned with accuracy.

As shown in Figure 9.2 below, Shoichi also shows a concern for accuracy, but this is embedded in a more complex set of actions and goals.

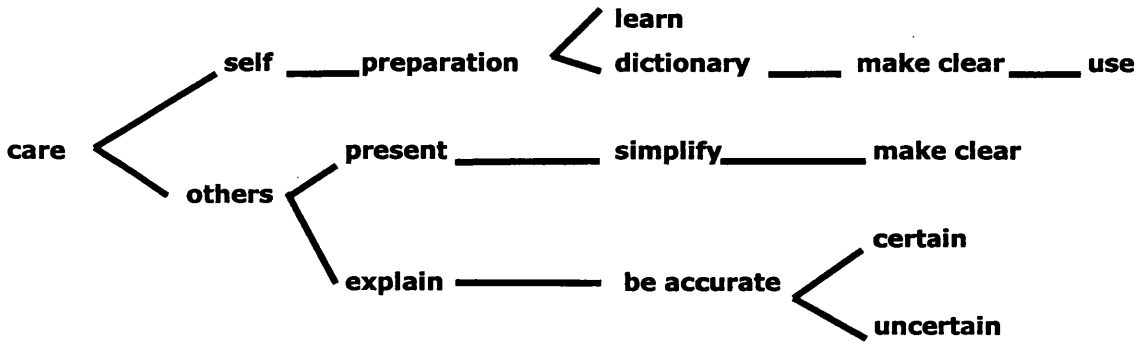

Figure 9.2 Care for overall process in Shoichi's account

Shoichi sees preparation as important for himself in that it enables him to learn and get ready for using English in class. He uses dictionaries to make things clear to himself before he communicates with others. His dictionary use also indicates that he prefers to create collocations from dictionary entries rather than retrieve them from text. In expressing his ideas to others, Shoichi is concerned with making things easy to understand ("simplify" & "make clear"), but he is also mindful of being accurate. This creates a tension for him because he wavers between feeling certain and uncertain about his own accuracy in English.

Emi's construct of care is similar in several ways to Shoichi's (see Figure 9.3 below). She shows concern for her own learning and for communicating her learning to others.

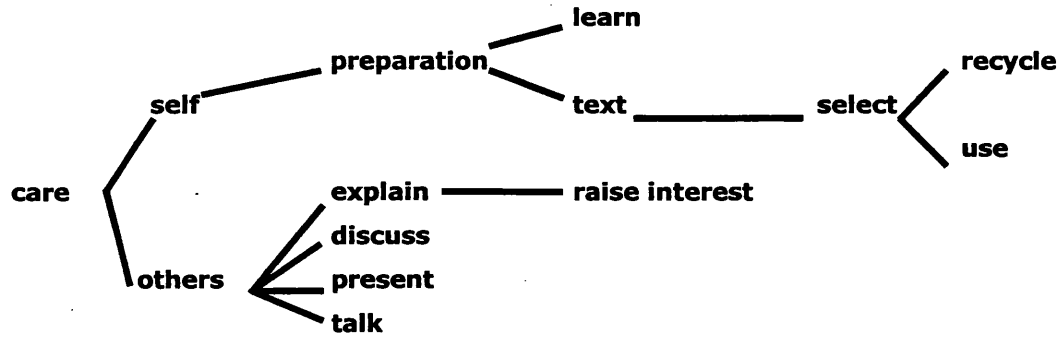

Figure 9.3 Care for overall process in Emi's account

What is different about Emi's construct is that she is more focused on selecting ideas and phrases from the texts that she is reading so that she can re-use these to herself and then be ready to use them in class. She is concerned with noting, rehearsing and reproducing conventional collocations rather than creating her own (unlike Shoichi). Emi is also concerned with raising the interest of others that she explains her research to. This affective side is positive and confident, whereas, with Shoichi, his (un)certainty about being accurate points to fluctuating confidence.

9.4.2 Engagement

This core category is defined as "involvement in trying to understand and to develop further understanding of a particular human rights issue through English." A content-based course where students choose their own issues to research in English may involve them in more than a series of discrete vocabulary and collocation learning tasks. Learners may begin to use English to understand the world they live in to develop their English and their understanding of human rights problems. The visual displays of each learner's construct of this category are presented in Appendix 9.2.

Kensuke's construct of engagement is minimally developed. His main interest is to learn fundamental words and facts about a particular issue. If Kensuke can understand fundamental words, then he has no need to develop his vocabulary further: "To some extent I understand fundamental words so I guess at first cycle I don't or couldn't need making vocabulary" (Kensuke, November). Shoichi, on the other hand, shows a degree of engagement that makes his research through English relevant to him in different domains.

His engagement is connected to his motivation to make his research clear to himself and others, which helps him improve his English ability. His research also helps him understand the world better. Emi's construct of engagement overlaps with Shoichi's. Her engagement enables her to make her learning relevant to her present life and her future. It also increases her motivation. Unlike Shoichi and Kensuke, Emi's engagement propels her to read more in English, develop her reading speed, and find more opportunities for reading. Emi's degree of engagement increases her chances of appropriating particular issue-specific vocabulary and collocations.

9.4.3 Reading

The core category of reading is defined as "using texts to find information and identify important ideas and vocabulary/collocations." Neither Kensuke nor Shoichi were found to develop a clear construct of reading texts as a means to a greater end. This is not to say that they didn't read, find information and record vocabulary/collocations. It is rather that they did not report this as a salient part of what they did. In Kensuke's case, his main focus was on either important words or new words. Shoichi, in contrast, saw texts mainly as containing important vocabulary that he would need to explain to himself and to others. Emi, however, had a quite complex construct of reading. She realised in the first cycle that she had no need to check every new word that she met; rather, she could be selective in deciding what to check. In making this decision, she freed up time for other purposes: "I gave so many new words here, but some are not so important for me—for me and also other classmates—to understand the point. So, I don't have to write down all the new words here. Instead of that, I can split my time to remember or connect with the context"

(Emi, November). Emi also valued repeated reading because this helped her learn more and came to see reading as a means to identifying important parts of a text and retrieving important words and their collocations.

9.4.4 Vocabulary selection

The core category of vocabulary selection is defined as "choosing vocabulary from text for further action." All three learners construct vocabulary selection as a decision about what to do with vocabulary that is either new or important for them. In Kensuke's case, there are few further actions that he takes (see Figure 9.4 below). He makes an effort to remember selected new vocabulary, but does not report any explicit strategies for doing so.

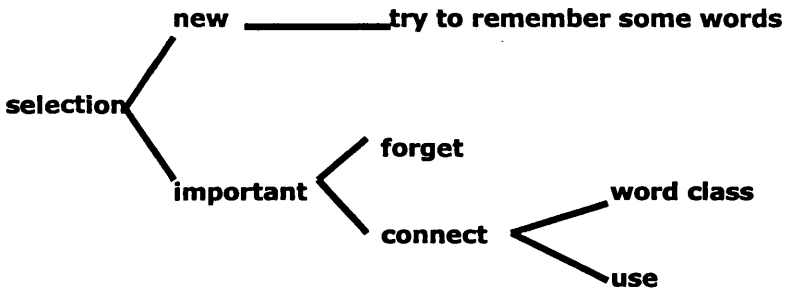

Figure 9.4 Vocabulary selection in Kensuke's account

As for important vocabulary, Kensuke either forgets or tries to connect (i.e., start to develop important collocations). According to the interview data, his collocation focus rests much more on formal word class accuracy (i.e., colligation) than meaning or possible lexical combinations.

During the interview period, Shoichi sees remembering all new words as important. At the same time, he comes to understand that using key vocabulary can help him

remember it. This has a positive affect on his motivation in that it helps him enjoy and feel confident about developing his ability to communicate his ideas to others. This is shown in Figure 9.5 below.

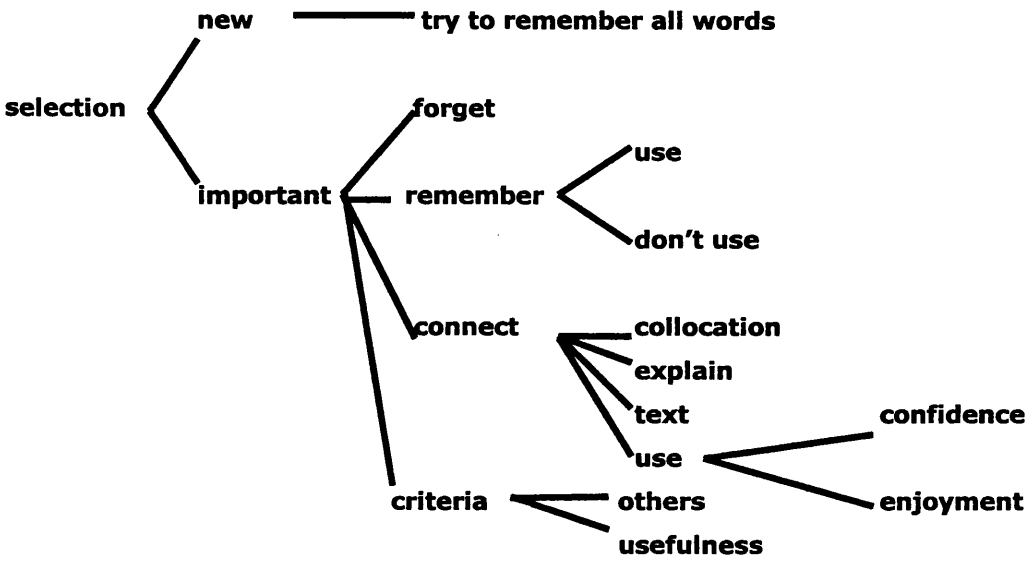

Figure 9.5 Vocabulary selection in Shoichi's account

Shoichi also views collocation as one way in which to make connections between words. In short, there is a positive edge to his construct of the processes of vocabulary selection. Collocation development is closely linked to increasing fluency and confidence.

As we have previously noted, Shoichi's and Emi's constructs tend to overlap. The main difference in Emi's construct of vocabulary selection is that time and practicality inform her decision to remember some rather than all new words (see Figure 9.6 below). Emi's further actions for connecting lexical items together are almost identical to Shoichi's, except in two aspects. First, she is explicitly concerned with meaning and is sensitive to making selections by importance and interest, as well as newness. The focus on interest shows a positive affective aspect to how she chooses vocabulary. Emi sees important and

interesting vocabulary as useful for helping her both research and communicate her ideas to others. Second, she is aware not just of forgetting, but also of what she forgets—which she identifies as formulaic phrases and infrequent words.

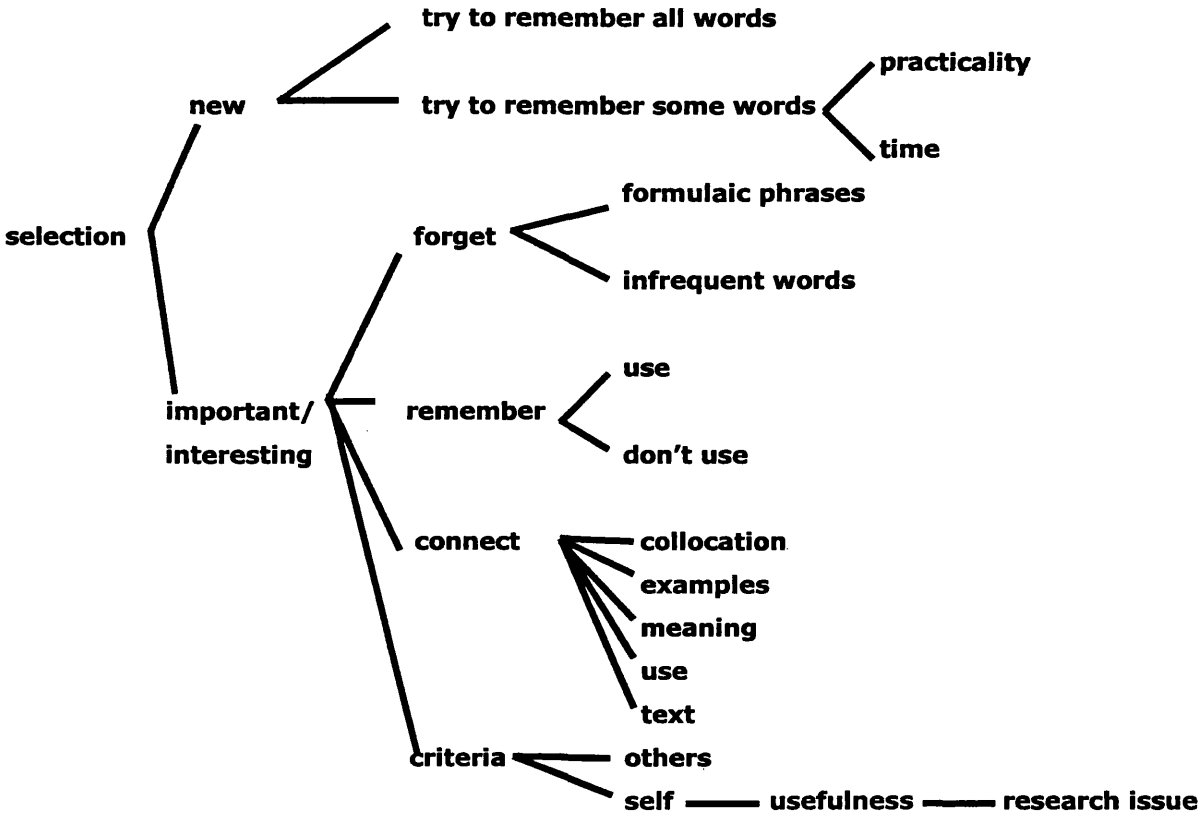

Figure 9.6 Vocabulary selection in Emi's account

9.4.5 Vocabulary connections

This core category is defined as "using different strategies to connect selected vocabulary to own knowledge and planned collocation use." As shown in Figure 9.7 below, Kensuke is mainly concerned with new and difficult vocabulary: "I make vocabulary notes like ensuring new words" (Kensuke, December) and "Normally simply I just check and

researched the meanings of difficult words" (Kensuke, January). By January, Kensuke shows a minimal ability to collocate key words on the issue of refugees, but is unable to connect his collocation learning with the way he makes notes: "I can't find any particular phrases in notes" (Kensuke, January). His approach is bound to text and dictionary at the time of meeting unknown words, and his overriding concern is with grammar rather than meaning. Grammar is "such a problem—I was careful and checked and tried to improve" (Kensuke, November). Kensuke has a very limited sense of paraphrase; he associates details, but does not elaborate the associative connections between them: "...work of refugees is for example groceries which are site of construction" (Kensuke, January).

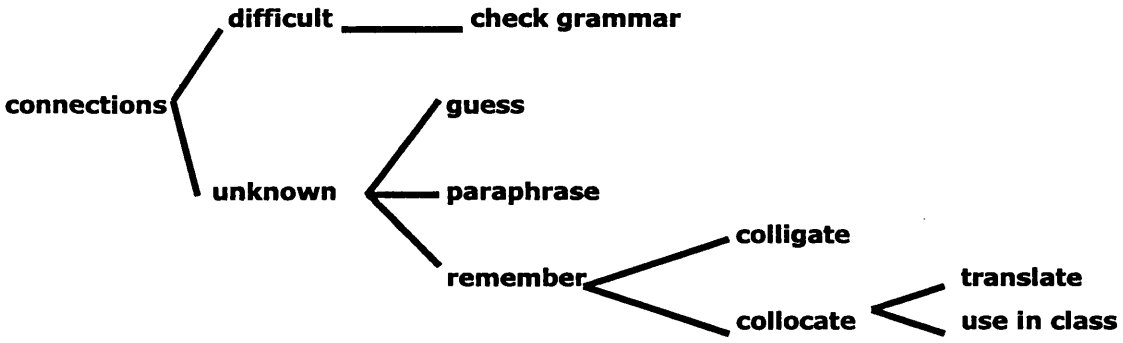

Figure 9.7 Vocabulary connections in Kensuke's account

In making vocabulary connections, Shoichi differs from Kensuke in that he begins to notice polar choices for himself in what he can do. He distinguishes difficult and easy vocabulary, as well as known and unknown. He also becomes concerned with creating links between words in the text or dictionary and collocations that he feels will be easy to remember: "In my mind with the dictionary for example **pact**...I can understand **arrangement** rather than **pact**...**Pact** means **arrangement**...**suicide pact** and **suicide**

arrangement." Shoichi also notices a difference between his everyday use of English and his explanations of his human rights research. Easy everyday collocations such as **give advice** are difficult for him to generate other collocations from: "**Advice** is very important in my life...but **give advice** I can't try to make other phrases" (Shoichi, January). On the other hand, he feels it important to simplify more specialized words like **insomnia** so that he can explain them in class clearly: "I try to transfer into easy or useful words in my mind so it's very easy to have presentation and explain" (Shoichi, November). An important part of 'creating phrases' lies in finding more frequent words to use for discussion, presentation and rehearsal (as shown in Figure 9.8 below).

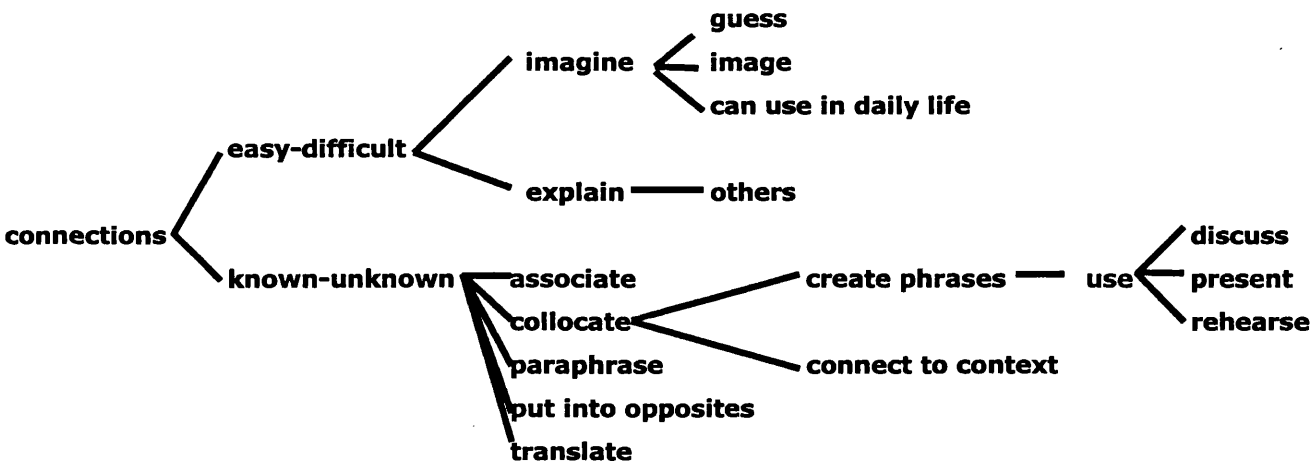

Figure 9.8 Vocabulary connections in Shoichi's account

As with previous constructs, Emi's construct for vocabulary connections overlaps with Shoichi's but is more elaborate (see Figure 9.9 below). One similarity is that Emi is concerned with creating meaningful connections in English to what she already knows and can use: "When I see **disreputable**, it is difficult to imagine the word. But when I see

respectable and **disreputable**, I can remember much more easily because I know the meaning of **respectable**" (Emi, November) and "When I see the new word like **combat**...I tried to write down the same meaning as **fight** so that it's easy to remember" (Emi, December). She differs from Shoichi in that she consciously translates into Japanese those words that she decides are unimportant, as well as focuses more on learning collocations in English for key words in texts on different human rights topics: "For the keywords, I could concentrate more on remembering the connection, the collocation. And many times I used them in discussion or presentation, and that kind of words are difficult to forget" (Emi, January). Paraphrasing and making collocations in English help Emi use and re-use important vocabulary, as much as using and re-using important vocabulary help her paraphrase and collocate in English.

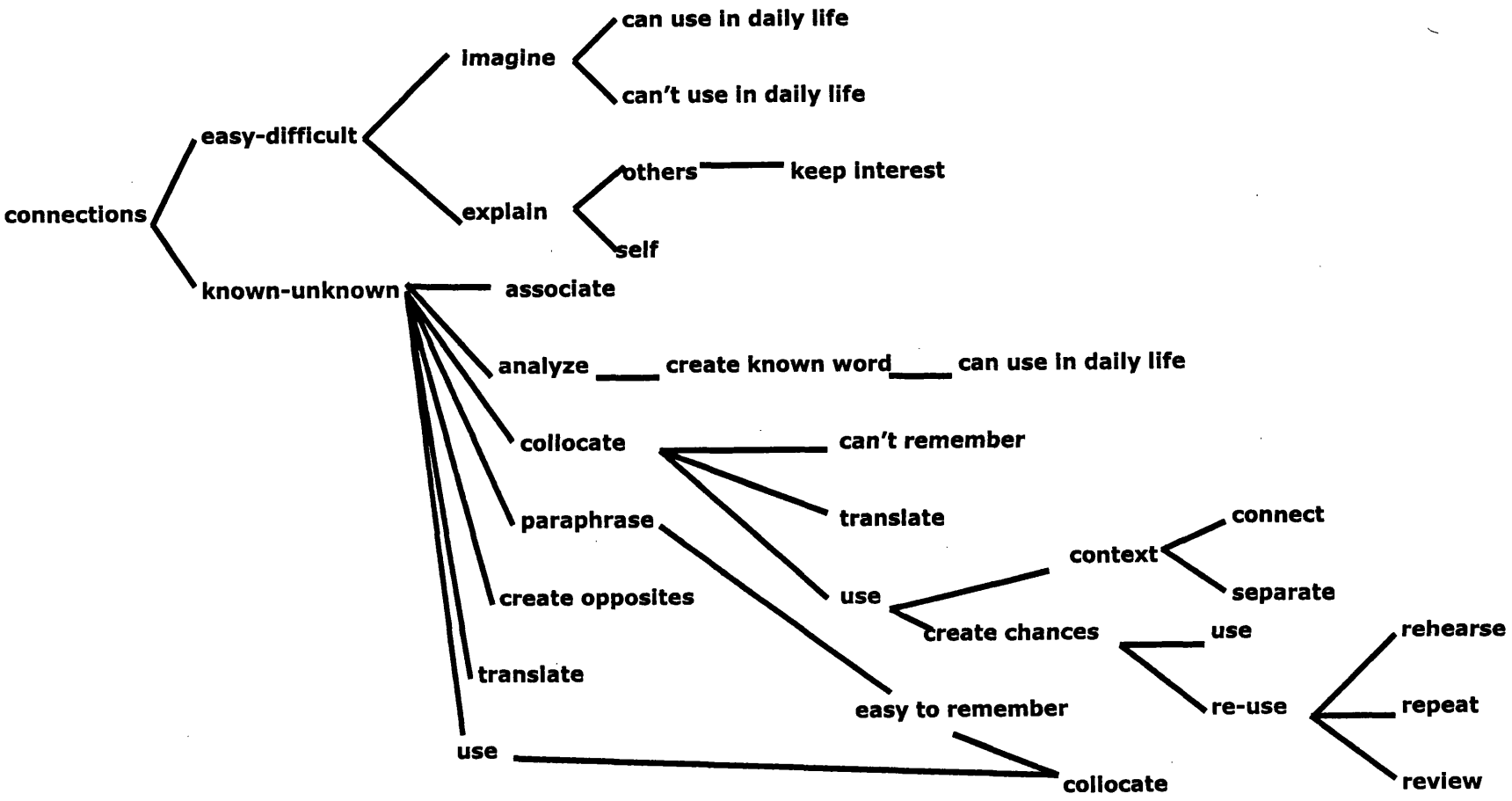

Figure 9.9 Vocabulary connections in Emi's account

9.4.6 Dictionary use

The core category of dictionary use is defined as "using dictionaries (English-Japanese, Japanese-English, monolingual English, monolingual English collocation) to take further actions with vocabulary/collocations." Kensuke's primary concern is with ensuring the meaning of "misunderstanding or unknown word" (Kensuke, November). He explains that he uses Japanese-English, then English-English and English collocation dictionaries, and that he makes few notes in the belief that "if those vocabulary [are] really necessary for me to use, I will remember those vocabulary" (Kensuke, November). There is very limited evidence of Kensuke recording and building particular collocations: "I was not perfect in that I didn't make any particular examples or phrases by using collocation dictionary" (Kensuke, January).

Shoichi is quite different in his use of dictionaries. He evaluates his previous dictionary learning as unreflective: "I have used dictionary unconsciously in this process" (Shoichi, November). For the first cycle, Shoichi explains how he often consults an English-Japanese dictionary, then a Japanese-English dictionary (and, if necessary, an English collocation dictionary) to create appropriate collocations: "I check the word in the English-Japanese dictionary. Next I make my mind clear, so then if I know the other words, I try the other words and check the collocation and then combine" (Shoichi, November). One example is **insomnia**. He finds 'fuminsho' in his EJ dictionary, then finds **sleeplessness** as the English translation for 'fuminsho' in his JE dictionary, together with the phrase **suffer from sleeplessness**; he uses this collocation to help him paraphrase **insomnia** in class. In the December interview, Shoichi reports greater use of the English collocation dictionary as he moves from focusing on important individual

words to noting and creating collocations for key words: "I know the meaning of **isolation**, so I tried to make it a phrase: **emotional isolation, isolation in their minds, social isolation, they suffer from isolation**" (Shoichi, December). In January, Shoichi explains that he now retrieves some key words with their collocations from text (e.g., **harassment violates women's rights, suffer harassment**, and **be subject to harassment**), as well as uses the collocation dictionary to find appropriate combinations for other important words (e.g., **give-get-offer compensation** and **achieve-make recovery**). He notes that his dictionary use is affected by the degree of certainty that he feels about different words, as well as by his perceived need to explain his research to others. Over the period of the interviews, Shoichi's increasingly more conscious use of dictionaries reflects his growing confidence in recognising, checking and using collocations.

Like Shoichi, Emi is initially uncritical of how she uses dictionaries in different ways: "Maybe there's no reason" (Emi, November). Another problem that she faces is noticing combinations with frequent words when she reads: "When I read the passage, I can understand it, so it is easy to pass" (Emi, December). However, in differentiating between unimportant and important vocabulary, she begins more and more to use, first, the English-English dictionary to develop her paraphrase ability and, second, the collocation dictionary to find and recognise useful collocations for key words: "When I read the article, I can...understand the sentence, so I know the keywords very well, but I don't know how it is commonly used, so the collocation dictionary helped me a lot" (Emi, January). Previous encounters with particular, useful words prime her use of the collocation dictionary:

"...**intimidate** can be used in many ways, because I have seen this word in other situations before. So I remember that, and it seems to be important to check the collocation dictionary" (Emi, December). She chooses **be easily intimidated**, but notices that it is still easier for her to remember the paraphrase of **be easily frightened** than the solitary collocation. Emi explains that she rejects uncommon or specialised words for collocation development (e.g., **strap**, **look-out**, and **amalgamation**), but still finds it difficult to imagine situations of use for particular words that she wishes to collocate.

9.4.7 Noticing

Noticing is the seventh of the eight key categories and is defined as "perceiving vocabulary/collocations in text and making informed judgements about what further actions to take." Kensuke's sense of noticing is restricted to checking unknown words: "I just check and researched the meanings of difficult words" (Kensuke, January). His concern with formal accuracy and difficult vocabulary short-circuit the development of his meta-cognitive decision-making.

Unlike Kensuke, Shoichi notices a fundamental shift in his approach to learning vocabulary in that he becomes more concerned with how to use phrases than with how to learn difficult words: "...to make certain phrases it's very possible to have discussion with confidence" (Shoichi, November). As we saw earlier, Shoichi realises that he can use different dictionaries to help him achieve this goal. This requires a very conscious effort which he finds somewhat contradictory. This is because he wants his English learning to be unconscious and 'natural'. By the end of the interview period, Shoichi reports that using key words and explaining his ideas to others have helped him to decrease his

dependence on Japanese and to increase his motivation for learning and using English collocations with greater fluency.

Emi's meta-cognitive decision-making is pervasive (see Figure 9.2D in Appendix 9.2 for a visual display of her construct of noticing). Like Shoichi, she sees her shift from checking all new words to focusing on important words as a fundamental development. She is concerned with building her phraseological competence with known vocabulary. Here, she makes a difference between vocabulary that she has met before and vocabulary that she has used before. If she notices that she has used particular vocabulary before, she feels less of a need to spend time on collocation building. With vocabulary that she has met but not used before (and which she decides is important), Emi decides to use the collocation dictionary and make notes to help her future output. She also reports making several other polar types of judgment about particular vocabulary items, namely difficult-easy, exact-partial and frequent-infrequent. Emi is also aware of how she uses different source websites for different purposes. Newspaper articles and BBC webpage texts on human rights are relatively easy for her to read because they contain very specific cases from different parts of the world; other websites such as the ILO, UNHCR and UNICEF contain more specialized and credible texts, with statistics and specialist explanations. In short, Emi becomes consistently pro-active and sophisticated in the decisions that she makes about how to learn English vocabulary.

9.4.8 Problematization of learning

This final core category is defined as "developing a sense of past English learning preferences, questioning them, and developing new ways of learning English according to

individual experience, awareness and goals." At the very end of the interview period, Kensuke reports that "this class influenced my studies of English" (Kensuke, January). He explains that he has become concerned with the structure of presentations rather than sentences or paragraphs and that he has started to use dictionaries to develop his way of learning vocabulary. It is, however, very difficult to ascertain whether he in fact developed specific new learning goals for himself.

By January, Kensuke appears to have changed his learning preferences very little from the beginning of the academic year. His preference remains mainly one of getting by on what he already knows. In an interview prior to November, Kensuke explained that, in high school, he had put himself under enormous stress to be successful in passing the English entrance exam for university; after that, he had lost interest in developing his English further, although he was happy to have the chance to talk with an English native speaker during the interview period. In short, Kensuke had no experience of learning English vocabulary other than by translation of individual difficult words for competitive exams. He remained resistant to changing a style of learning in which he had invested so much in order to get into the Faculty of Law in the first place.

Shoichi noticed different aspects of his learning preferences and tried to change some of them. He reports that "in the first semester when I found only the meaning [of new words], I stopped to learn" (Shoichi, November). In his previous English learning, Shoichi felt that he had been neither good nor confident at using English. His basic change in focus from "difficult" to "useful" was critical in helping him manage phrase-based vocabulary learning and develop his fluency in English. According to Shoichi, he had

been learning English for 11 years, but neither he nor his other teachers had ever had time or opportunity to discuss different ways of learning vocabulary. During the year, Shoichi regained his confidence and became highly motivated about developing his ways of learning English vocabulary and collocations.

One of the singular influences on Emi's development of her English learning was that she completed an internship at the United Nations in Geneva during the summer. In Geneva, she read many documents in English and participated in international youth forums on human rights: "I could read many reports in a short time" (Emi, November). This affected her learning preferences once she returned to Japan: "After the summer, I took care of other classmates who is listening to my report and I thought much more about them" (Emi, November). This concern with audience helped her organise her ideas more clearly and motivated her to rehearse before class. In the December interview, Emi reported how she was using texts rather than dictionaries as her primary source of collocation input. This, in turn, led her to develop her way of making notes and to use dictionaries more strategically: "If I focus on key words more, it helps me to use it with confidence" (Emi, December). She also explained that until recently she had been concerned with learning difficult vocabulary in bilingual lists; she had felt satisfied "if I could know the meaning in Japanese" and "translate it from English to Japanese" (Emi, December). The other important factor that affected the development of her English learning style was her interest in human rights. Emi was clear that she wanted to study human rights issues further in graduate school, so her future goals strongly motivated her to find better ways of studying and using English vocabulary and collocations.

9.4.9 Emerging theorization of collocation development

The different personal constructs that Emi, Kensuke and Shoichi have for the eight key categories can be briefly summarised as a series of critical tensions. These key tensions are shown in Figure 9.10 below.

own linguistic knowledge	⟷	communication with others
uncertainty	⟷	confidence
lexical accuracy	⟷	lexical fluency
lexical form	⟷	lexical meaning
lexical difficulty	⟷	lexical usefulness
lexical infrequency	⟷	lexical frequency
memorization	⟷	re-use
reading for new words	⟷	reading for ideas
remember in Japanese	⟷	connect in English
translate	⟷	paraphrase
colligate	⟷	collocate
individual words	⟷	phrases
error avoidance	⟷	explanation
random dictionary use	⟷	strategic dictionary use
single reading	⟷	repeated reading
engrained habits	⟷	conscious goals
resistance	⟷	reflection

Figure 9.10 Tensions emerging in a theorisation of L2 collocation knowledge development

The tensions inform the different approaches to building L2 collocation knowledge that each individual takes. They are furthermore relative to each individual's history, awareness of their actions in the present, and future goals. The more particular aspects of an individual's learning are located to the left, the less likely they are to be successful in developing L2 collocation knowledge. At the same time, the more a learner moves

particular features of their learning to the right, the more likely it is that they will develop their L2 collocation ability.

9.5 Discussion

This longitudinal case-study set out to explore the learner's side of the story in developing L2 collocation knowledge. Through interviewing three learners, our goal was to uncover a set of personal constructs informing their English vocabulary and collocation learning. The constructs that emerged were:

- care
- dictionary use
- engagement
- noticing
- problematizing learning
- reading
- vocabulary connections
- vocabulary selection.

For each construct, we examined the learners' own accounts to identify similarities and differences. This allowed us to draw out a set of tensions affecting the local and contextualised development of L2 collocation knowledge. In this section, we will consider particular aspects of those tensions in greater detail.

It is clear that Kensuke's overall construct of vocabulary and collocation learning is, in many respects, different from Shoichi's and Emi's. What largely determines that strong sense of difference is his resistance to breaking free of his previous test- and translation-based learning experiences. The education system in Japan demands such preparation for the university entrance exam, and any account of English learning at the tertiary level in Japan will be affected by such institutionalized (and internalized) pressure. Kensuke continued to prefer formal linguistic knowledge to using English for learning about the world. He was disengaged from being able to move towards a greater concern with

fluency and, by implication, sustainable English collocation development. We saw in Chapter 7 that the quantitative results showed that Kensuke had a large L2 vocabulary like Emi, but, unlike Emi, he flat-lined in the development of his English collocation knowledge. The interviews with Kensuke now allow us to understand very clearly why he did not develop his collocation scores: He was not particularly interested in making an effort to do so and relied on his existing extensive knowledge of single English words. Through the year, Kensuke was mostly concerned with decoding new and difficult vocabulary into Japanese. In this way, he maintained control but did not make much observable progress.

The findings from this chapter indicate that Shoichi and Emi also faced the challenge of breaking free of their ingrained English vocabulary habits. Like Kensuke, they too had become accustomed to learning English through translation and focusing on the difficult and new rather than the useful and already known. In Emi's case, one of her previous English classes at university had been discussion-focused, which had helped her develop her confidence in using English. However, her previous English learning experiences had never involved her in considering different ways of learning English vocabulary. What helped Emi become more strategic was her experience as an intern at the UN in Geneva in the summer before the interview period presented in this chapter. The internship enhanced her confidence and helped her increase her reading speed and fluency. It also provided her with a positive foundation for re-considering how to develop and sustain her English collocation learning in the second semester. These general comments about Emi's performance offer a plausible explanation for the constant increase in her collocation

scores that we noted in Chapter 7.

In contrast, the quantitative analysis of Shoichi's collocation knowledge in Chapter 7 showed that he fluctuated markedly, but overall made clear progress. The word association and collocation scores also demonstrated that he drew on frequent vocabulary more than either Kensuke or Emi. The interview data from the present study allow us to understand why that was the case. The different accounts that Shoichi puts forward show that, in the first semester, his confidence was initially low for using English to explain his ideas and for changing his way of learning vocabulary. After some time, he re-considered his existing learning habits and began to focus more and more on making his research clear—first, to himself, and, second, to others. This pushed him to using different dictionaries in distinctly creative ways to paraphrase specialized vocabulary and build useful collocations that others would be able to understand. Once he realised that such preparation helped him increase his fluency, his confidence grew, and, by the end of the interview period, he was retrieving key word collocations from texts and not just creating his own from using dictionaries. In this sense, Shoichi made strong progress in developing his L2 collocation knowledge.

Part of the overlap between Shoichi and Emi is that, in their own ways, they came to see the parallel importance of paraphrase and collocation building for developing their English vocabulary and fluency. This is perhaps the most interesting result from the interviews. It is clear from the data that both learners see paraphrase as a means of using more frequent vocabulary that they know well to explain less frequent words that others will have trouble understanding. Paraphrase, then, becomes a space in which they perform

two important lexical operations to foster their L2 collocation knowledge. On the one hand, they use paraphrase to make interconnections between frequent and infrequent vocabulary—to re-activate and connect what they already know so that they can explain something that is new to them. On the other hand, paraphrase creates the need for concise and transparent collocations that they can confidently and comfortably use to explain their ideas to others. This interpretation places paraphrase as the critical platform for L2 collocation development.

This insight is not only interesting, but also somewhat surprising if we recall that collocation knowledge is often seen as an advanced level of lexical competence for individual words. Such a view is common within depth of knowledge views of L2 collocation knowledge: Collocation is the icing on top of the lexical cake. However, the evidence from the interviews suggests otherwise. We may now advance a counterview to depth of knowledge views of L2 collocation knowledge. An alternative view is that paraphrase acts as a priming mechanism for developing conventionalised associative links between L2 lexical items. Collocation is more than mere association. L2 word association knowledge tends to be highly idiosyncratic, but collocations are expected and conventionalised links between lexical items to express ideas that particular speech communities arbitrarily agree on. Using paraphrase helps an individual move from idiosyncratic to conventionalised associative, i.e., collocation, links in their L2 lexicon.

9.6 Conclusion

This chapter was concerned with how learners come to know and use collocations in

English. In this study, we investigated in detail how individual learners organize, understand and interpret their own development of lexical and collocation knowledge over time. This inquiry followed a grounded theory analysis and examined not just linguistic, but also cognitive and meta-cognitive dimensions of L2 collocation knowledge. The analysis allowed us to identify eight key categories and to develop our understanding of these categories through close examination of three learners' personal constructs. Several interesting differences and similarities emerged, which let us triangulate in qualitative terms some of the major findings from the parallel longitudinal quantitative study presented in Chapter 7. We were also able to identify a critical role for paraphrase in the development of L2 collocation knowledge. The findings indicated that paraphrase has a key function in interconnecting both frequent-infrequent lexis and idiosyncratic-conventionalized associative knowledge.

The conclusion to this chapter brings us to the end of the experimental work for this thesis. In the next chapter, we will briefly review the major insights from all the different experiments before we engage in an extended discussion of the development of L2 collocation knowledge.

Chapter 10
Discussion

10.0 Overview

In this chapter, I would like to look back at the research narrative as a whole and to draw out salient issues for discussion. I will begin by briefly recalling the different areas of potential inquiry for L2 collocation research that I presented at the end of Chapter 2. This will enable us to re-direct our attention to some of the gaps that I have attempted to address in my own research. I will then continue by recapping the main insights from each experiment and focus these insights towards a set of issues for discussion. After discussion of these issues, I will put forward a model of L2 collocation development. Finally, I will conclude this chapter with some further questions to consider in researching L2 collocation development.

10.1 Gaps in exploring L2 collocation knowledge

The literature review in Chapter 2 let us identify various gaps in previous studies of L2 collocation knowledge. We saw that previous research was largely concerned with measuring knowledge of adjacent 2-word collocation types. This focus on type was accompanied by a concern with mis-collocations. I concluded that it would be helpful to look not just at collocation types, but also at the collocation spans of different lexical items. I also put forward the possibility of developing decontextualized measures of L2 collocation knowledge to help us move beyond a preoccupation with collocation error. We furthermore found that previous studies generally used measures that tested relatively few items, so another area of concern was to develop measures that included a relatively large number of items. A further problem with previous research was that

most studies used intermediate to advanced learners, so they offered a rather restricted view of L2 collocation proficiency. We also noted that it would be helpful to compare L2 collocation knowledge to an independent measure of L2 proficiency. The final gap that we identified was that previous research had generally failed to consider in any great detail how learners themselves go about developing their L2 collocation knowledge. In other words, it would be useful to investigate not just what learners know but also to find out how they come to know what they do. All in all, the gaps in the previous research pointed to several interesting areas of potential inquiry.

10.2 Summary of insights

From reviewing previous studies, I decided to investigate L2 collocation knowledge and production around three overarching questions:

- What collocations do learners recognize?
- What collocations do learners produce?
- How do learners go about learning collocations?

These were the general points of departure from which I generated more specific areas of inquiry to explore some of the issues that previous research had pointed to. Those three questions now offer us a useful way to pull out key insights from the research narrative of my own experimental work. So, under the general headings of collocation recognition, production and process, I will briefly re-frame the main insights from the different experiments that I have presented in the previous seven chapters. (For a sequential presentation of insights by chapter, see Table 10.1 on the next two pages.)

Table 10.1 Summary of insights from each experimental chapter

Chapter	Focus	Insights
3	Corpus analysis of high & low frequency verbs and nouns and their collocates in written productions	• The productive use of verbs is limited; frequent verbs are the most widely used. • The collocation spans of AWL verbs tend to be lexically rich and include noun phrases where AWL nouns are combined with frequent adjectivals. • The collocation spans of frequent verbs tend to be lexically poor and close to spoken English.
4	Recognition of AWL verbs using Nation's cluster format as the test instrument	• Frequent AWL verbs are the most strongly recognized. • Partial knowledge of lexical items is critical. • Learners exploit three main types of lexical knowledge: paraphrase, associative networks, and combinations of multiple lexical strategies. • Problems with the cluster format point to the need for a different instrument.
5	Recognition of frequent & AWL verbs using a knowledge-state ranking instrument	• The knowledge-state ranking instrument performs well. • Very frequent verbs (i.e, K1 items) are most known. • Verb/noun homonymy may influence learner recognition of lexical verbs.
6	Recognition of verb + noun collocations using a knowledge-state ranking instrument	• Noun recognition is stronger than verb recognition. • Noun recognition and verb recognition correlate significantly with general English proficiency. • Collocation recognition does not correlate significantly with general English proficiency. • Knowledge of *individual* verbs and nouns shows a significant relationship to knowledge of V+N collocations for particular states of knowledge.

7	Piloting of a new instrument through a qualitative case study to examine word association and collocation production longitudinally	Word associations: • Shoichi shows greater use of frequent vocabulary in word associations. • Kensuke produces different word associations more. • Emi repeats the same word associations more. Collocations: • Shoichi produces frequent collocates and inappropriate collocates more. • Kensuke produces different frequent and infrequent collocates more. • Emi repeats the same collocates more.
8	Development of a test of collocation production with 30 frequent nouns and examination of frequent and infrequent collocates	• More collocationally competent learners produce significantly more infrequent *and* frequent collocates. • For low proficiency learners, adjectival links are the dominant type of collocation. • For high proficiency learners, adjectival links remain frequent, but are also complemented by nominal and verbal, as well as nominal-verbal links. • Homonymous collocates are an important difference between high and low proficiency.
9	Exploration of learners' individual constructs of vocabulary and collocation learning & use	Learners differ in the elaboration of personal constructs for vocabulary and collocation learning. Collocation learning requires a number of conditions to be met, in particular: • breaking free of exam-based grammar-translation learning • distinguishing between new vocabulary and important vocabulary • balancing knowing with using • using paraphrase as a means to reach the edge of the ZPD where collocation use is most likely to occur.

10.2.1 Insights concerning collocation recognition

In three experiments (Chapters 4-6), I sought to understand second language collocation recognition. Despite Chapter 6 being the only episode directly concerned with testing V+N collocation recognition, the other two chapters—Chapters 4 and 5—involved preparatory work that allowed for the design of a robust knowledge state ranking instrument. The key insights for collocation recognition can be reviewed in relation to:

- frequency
- word class
- partial knowledge
- general English proficiency.

I will briefly comment on each of these areas so that they can be elaborated as themes in the later discussion.

With individual lexical verbs, frequency was found to be a critical factor in recognition, whether these verbs were from K1-K2 ranges or the AWL. Recognition of individual frequent core and AWL lexical verbs showed a significant correlation with English proficiency because learners varied more in their recognition of lexical verbs than the nouns that they were collocated with. Noun recognition also correlated with general English proficiency, but it was so strong that zero variance could be observed in most cases. Thus, we may infer that noun recognition has a primary effect on V+N collocation recognition. By the same token, verb recognition may be thought to have a secondary effect on V+N collocation recognition. The insights from the suite of recognition tests suggested that nouns may be very important as stimulus items in productive tests of collocation knowledge. Nouns would, it was assumed, act as reliable anchors to which other adjectival or verbal lexical items could be collocationally linked, whether appropriately or not.

The think-aloud protocols about the academic verb test presented in Chapter 4 showed that partial lexical knowledge is a critical factor in learner recognition of individual lexical verbs. We found that learners exploit three main types of partial lexical knowledge, namely paraphrase, associative networks and combinations of multiple lexical strategies. This let us see that viewing the development of vocabulary knowledge as a uniform incremental process from known to unknown is mistaken. That is, learners do not acquire increasing knowledge of a lexical item in linear stages from formal to collocational. On the contrary, once learners have established a certain degree of recognition, they are focused on connecting partially known lexical items by meaning. Although collocation knowledge was rarely used in the think-aloud protocols, we found that learners often tried either to create loose idiosyncratic semantic links (i.e., use personal associative knowledge) or to produce semantic extensions through what they knew more certainly (i.e., use paraphrase). We may assume then that associative knowledge and paraphrase can be more easily used than collocation knowledge on the periphery of lexical knowledge. The insights from the think-aloud interviews thus raise interesting questions for discussion about the organization of the L2 mental lexicon when learners are working at the edges of their knowledge. This in turn may allow us to interpret the relationship between different forms of lexical knowledge quite differently from the incremental linear view that Nation's taxonomy of discrete types of lexical knowledge suggests (Nation, 2001, p.347).

As mentioned above, we found that discrete types of lexical knowledge—here lexical verb knowledge and noun knowledge—correlated significantly with general English proficiency; we were also able to establish that knowledge of individual verbs and nouns shows a significant relationship to knowledge of V+N collocations. Yet,

these insights were tempered by the finding that collocation recognition did not significantly correlate with general English proficiency. We had evidence that, although learners showed strong knowledge of individual lexical items, they did not translate this into collocation proficiency. Again, these insights point us towards discussing organizational aspects of the L2 lexicon and considering why collocation knowledge lags behind other types of lexical knowledge. Connected to this is the question of whether we should accept the claim that L2 collocation knowledge is basically an advanced form of lexical knowledge (as is often argued by researchers in the field). On the contrary, are there grounds for claiming that collocation knowledge is a more pervasive type of ability that may operate at earlier stages of L2 lexical development? In sum, the central issues for discussion that emerge from the research into collocation recognition are:

- the centrality of noun knowledge in the development of L2 collocation knowledge and the potential consequences that this has for understanding the development of L2 collocation use;
- the role of associative knowledge and paraphrase in running ahead of collocation knowledge and their potential priming effect on the development of L2 collocation knowledge and use;
- the positioning and nature of L2 collocation knowledge in an overall understanding of how learners develop their L2 mental lexicons.

Having drawn out these three issues for discussion, I will next consider what can be highlighted from the collocation production experiments.

10.2.2 Insights concerning collocation production

Three experiments (Chapters 3, 7 & 8) explored second language collocation production. The first was corpus-based and used written data, and the other two experiments made

use of noun prompts for eliciting productive collocation knowledge. The main insights from all of these inquiries can be related to the following areas:

- lexicality
- delexicalisation
- links in associative and collocation networks
- collocate frequency
- word class and English proficiency.

I will briefly look at each of these areas so that we can again draw out themes for later discussion.

The written corpus analysis in Chapter 3 showed that the collocation spans of AWL verbs tended to be lexicalized in that they involved AWL nouns which were pre- or post-modified in different ways. On the other hand, the collocation spans of frequent verbs were lexically empty in that they often consisted of highly frequent pronominal collocates. Lexicality here was marked by the use of an infrequent noun with an infrequent verb, where the noun may have also been modified by a frequent adjectival. From this analysis emerges the question of how we might characterize the way in which language users combine frequent and infrequent lexical items as collocations. The collocation environments that we have just characterized as lexically empty suggest that delexicalisation is one important aspect of L2 collocation development. How should we interpret more proficient learners' ability to combine and delexicalize items at the same time? Is the corollary that learners with weak collocation knowledge tend to use fully lexicalized items in collocations, and if so, why? This is another issue for discussion.

The experimental work in Chapters 7 & 8 led to several insights about links in collocation networks. In the longitudinal case study of Chapter 7, we also considered links in word associations over time. The case study suggested that low and high

proficiency learners may differ both in the kind of links that they make and in the frequency with which they re-produce the same links, whether it was a question of word associations or collocations. We had some indication that low proficiency learners may activate different links but tend not to repeat them, whereas high proficiency users can produce multiple links and repeat them. In Chapter 8, another difference that we observed concerned the frequency profiles of collocates produced by more proficient and less proficient learners. More proficient learners produced frequent and appropriate collocates more, as well as infrequent and appropriate collocates more. We also found significant differences in the types of word class that the two groups of learners produced as collocates. Less proficient learners' collocate productions were adjectival in contrast to the greater variety of word class forms—adjectival, nominal, verbal, and nominal-verbal—that the high group supplied. Earlier, we mentioned that general English ability did not correlate significantly with collocation recognition; yet, for collocation production, there was a significant correlation. So, two issues emerge here for discussion: first, why word class variation should accompany collocation development, and second, what might plausibly explain the connection between proficiency and collocation production. These issues also raise interesting questions about how we may define L2 collocation ability within particular contexts of restricted exposure to English—and whether it makes sense to define such an ability in exclusive relation to a distant native-speaker standard. All in all, the main issues from the research into collocation production are:

- the combination of frequent and infrequent items in L2 collocations
- the tensions between delexicalisation and lexicality in L2 collocations
- the interconnectivity of collocated items in the L2 mental lexicon
- specific contextual characteristics of L2 collocation development and ability.

In the final part of this section, we will consider insights about individual collocation process that are relevant to further discussion.

10.2.3 Insights concerning individual collocation process

The final experimental chapter (Chapter 9) differed from the previous quantitative research. It involved a longitudinal qualitative study of three learners' constructs of their vocabulary and collocation learning within a content-based English course. The grounded theory analysis approach allowed us to establish several key conceptual categories to understand how learners vary in their elaboration of personal constructs for vocabulary and collocation learning. Among the different questions that arose from the analysis, three in particular can be considered relevant to further discussion. These are:

- In what ways can we assume there is a connection between what learners see as new/difficult and important vocabulary and the distinction between frequent and infrequent lexis that we highlighted in the quantitative research?
- Do the quantitatively established insights that we have claimed for collocation recognition and production hold for what learners themselves see as part of their own learning processes? If they do, then in what ways? If they do not, then how can we explain the gap?
- Do the quantitative insights about associative and collocation interconnectivity stay plausible when we consider the qualitative insights that we reached—particularly for the tension between knowledge and use, and the role of paraphrase as a means of priming L2 collocation use and development?

These are the central questions for discussion that the last experimental chapter takes us towards.

10.3 Issues for discussion

We have identified several issues for discussion under the separate rubrics of collocation recognition, collocation production, and collocation process. We can now bring these questions together and re-categorize them into three areas of focus, namely:

- exploring dimensions of collocation knowledge and use
- linking lexical knowledge and use with collocation development and context
- proposing an emerging model of L2 collocation development.

Table 10.2 below shows the particular issues that pertain to each area of focus.

Table 10.2 Issues for further discussion

Area of focus	Particular question
Exploring dimensions of collocation knowledge and use	• the tension between delexicalisation and lexicality in L2 collocation use • the combination of frequent and infrequent lexical items in L2 collocation use • the centrality of noun knowledge in L2 collocation knowledge and implications for L2 collocation development
Linking lexical knowledge and use with collocation development and context	• situated networks and the issue of native-like collocation ability • the priming role of associative knowledge and paraphrase for L2 collocation development • the interconnectivity of collocated items in the L2 mental lexicon • the role of context in L2 collocation development
Proposing an emerging model of L2 collocation development	• characteristics of the model • connections between psycholinguistic and contextual dimensions of L2 collocation development

I will continue by exploring these three areas in more detail.

10.4 Exploring dimensions of collocation knowledge and use

There are two dimensions that I will discuss here: delexicalisation and lexical frequency. Both dimensions influence L2 collocation development, in which nominal groups act as a central anchor, the third topic of discussion in this section.

10.4.1 The tension between delexicalisation and lexicality in L2 collocation use

I will first consider a narrow understanding of delexicalisation based on V+N combinations, before exploring how delexicalisation can be understood more broadly as one of two important dimensions of collocation use. I will also examine how learners necessarily resist delexicalisation as part of the process in how they develop their L2 collocation knowledge.

Delexicalisation has most often been used for describing a particular group of V+N combinations in English where the verbal component carries little meaning (Sinclair, 1990a; Chi Man-lai, Wong Pui-yiu, & Wong Chau-ping, 1994; Allan, 1995; Hunston & Francis, 2001; Altenberg & Granger, 2001). Typical examples are phrases such as **have a shower** and **take a photograph**. Wierzbicka (1982), Stein (1991) and Stein and Quirk (1991) show that verbs such as **give**, **have** and **take** have particular semantic distinctions in V+N combinations. However, Allan (1995, 1998) finds that, despite **give**, **have** and **make** largely keeping their full "intuitive meaning," there are many instances where they do not. Indeed, the dominant view is that delexical verbs, i.e., **give**, **have**, **make**, **take**, **do**, **hold**, **keep** and **set** (Sinclair, 1990a), are empty or general in meaning and thus desemanticized (Sinclair, 1991, p.112; Allan, 1995, p.4; Stubbs, 2001, p.32). The desemanticization in such V+N combinations results in a focus on a completed event rather than a particular action. For example,

Comis took a photograph of her points to the completed event, whereas **They photographed the pigeons in Trafalgar Square** focuses on the action of photographing (Sinclair, 1990a, p.147). The completed event involves a concrete and specific dynamic action, whereas the full lexical verb tends to convey a stative sense (cf., Tanabe, 1999, p.130, in her discussion of V+N combinations in the 15th century *Paston Letters* from a period just before the proliferation of these 'general' verb combinations in Modern English). These V+N collocations are intriguing in that they appear to highlight an extreme case of semantic loss in the verbal component (or 'semantic bleaching,' as Phillip (2004) describes it).

Yet, it is not just a small set of V+N collocations that feature delexicalisation. It may also be argued that collocation in general leads to desemanticization in (at least one of) the words that are collocated. Tognini-Bonelli, for example, defines delexicalisation in the following terms:

> Delexicalisation here can be defined as a collocational relationship where one word loses most or all of its specific semantic content, while its collocational partner appears to keep most of its characteristic meaning. (Tognini-Bonelli, 2001, p.116)

She adds that delexicalisation "is mainly associated with words that are frequent" (p.117). Sinclair also argues that delexicalization is a pervasive feature of collocation and comments at length about this with the example of *dark night*:

> Collocation...restricts the meaning of the words involved, it does not enhance it. Now if collocation is a reduction in meaning, then *dark night* actually is not one of the meanings of *night* but a reduction of the meaning of *night*. This cannot be the answer. But it took a long time to give up the traditional concept of the word as unit of meaning. Once we accept that words can be co-selected, not chosen always one at a time, then there is no longer a problem with *dark night*; *night* does not distinguish one of the meanings of *dark*, nor does *dark* distinguish one of the meanings of *night*. The phrase *dark night* has its own meaning...the adjective *dark* is reinforcing the dark element already in *night*. (Teubert, 2004, *Interview with John Sinclair*, in Krishnamurty (ed.), p.xxi)

In other words, collocation involves semantic loss so that a particular pragmatic focus can be achieved with the collocation phrase. A collocation does not involve a doubling of the separate and complete semantic loadings of either element, but a reduction of meaning associated with (at least) one of the elements, as well as the creation of "a sense that is not the simple combination of the sense of each of the words" (Sinclair, 1991, p.104).

Once we place delexicalisation/desemanticization as a central process of collocation in general, we may see L2 collocation development in a different light. Sinclair frames the issue in the following terms:

> Many learners avoid the common words as much as possible, and especially where they make up the idiomatic phrases. Instead of using them, they rely on larger, rarer, and clumsier words which make their language sound stilted and awkward. This is certainly not their fault, nor is it the fault of the teachers, who can only work within the kind of language descriptions that are available. (Sinclair, 1991, p.79)

Sinclair indicates that the solution to the problem is the lack of appropriate descriptions of collocation patterns that need to be learnt. Yet, this externalizes the issue and directs our attention away from looking more closely at what is happening internally within the L2 mental lexicon. The learner's will to use 'larger, rarer, and clumsier words' may in fact be an important stage in L2 collocation development—something that necessarily happens rather than something that could be avoided under different circumstances. A specific example can help us develop this argument further.

A common V+N (mis-)collocation that Japanese learners produce is **improve the problem**. We would normally expect **improve the situation** or **solve the problem**, so we might characterize the learner collocation as a blend of two conventional alternatives. Pragmatically, however, it could be argued that **improve the problem** more or less gets across the meaning of 'taking action to deal with a

difficult situation.' It is a somewhat clumsy formulation, but it is sufficiently understandable and quite characteristic of L2 use at a particular stage of development. This stage can be defined as **lexicality in collocation use**, where lexical items retain their individual core semantic loading as they are combined.

If lexical items remain fully semanticized, they have a different semantic value or strength from items that lose all or part of their semantic content. What might happen, then, if we assign numerical values to represent such differences in semantic value? For the sake of discussion, we may give each lexical item in the collocation **improve the problem** a (complete) semantic loading value of .5, so the combination be described as having a loading of 1. Following Sinclair's claim that collocation restricts the meaning of a collocation, an appropriate collocation such as **solve the problem** will have a semantic loading <1. Although the precise loading is unimportant here, we can, for the sake of argument, make it .75 to distinguish it from the loading of 1 that we have just given to **improve the problem**. In our hypothetical example, we have now identified the semantic loss as .25 for the conventional (and partially desemanticized) collocation phrase, and as 0 for the learner (and fully semanticized) collocation. Table 10.3 below summarizes the argument so far.

Table 10.3 **Delexicalization and lexicality: a hypothetical example**

Collocation type	Verb	Noun Phrase	Loading
Conventional (+ delexicalization)	solve (.25)	the problem (.5)	.75
Learner (+ lexicality)	improve (.5)	the problem (.5)	1.0

If we apply this view of delexicalisation and lexicality to vocabularies rather than individual collocations, we can begin to draw out some interesting differences in the

development of collocation knowledge. We could, for example, take vocabularies ranging in size from 100 words to 1000 words and stipulate that in the conventional vocabularies each item collocates with one other, and that each pair of collocated words is separate from any other. This would lead to 50 conventionalized collocations per 100 words, and a collocation loading for each vocabulary as shown in Table 10.4:

Table 10.4 **Conventionalized collocation knowledge and vocabulary size: a hypothetical example**

Vocabulary size	Collocation links	Lexicality loading
100	50	75
200	100	150
300	150	225
...	...	...
1000	500	750

In such vocabularies, the lexicality loading is always less than the size of the vocabulary. On the other hand, for the learner vocabularies, we might stipulate that such vocabularies start with full lexicality so that the size of the vocabulary and the lexicality loading are initially the same. This would represent zero conventional collocation knowledge. We could then gradually increase the number of paired links (i.e., conventional collocations) so that small developments in collocation knowledge are included, as in Table 10.5 below:

Table 10.5 **Learner collocation knowledge and vocabulary size: a hypothetical example**

Vocabulary size	Collocation links	Lexicality loading
100	0	100
200	0	200
...	...	...
500	10	495
...	...	...
1000	100	950

This basic model of collocation knowledge uses only discrete pairings between words, all of which only link once with another word. We can imagine that the model becomes more complex and interesting when multiple links between different words are taken account of—where delexicalisation is seen as a gathering process of interlinkage across vocabularies. We could also take such interlinkage to the extreme and make calculations for a completely delexicalized vocabulary where all items are linked to all other items. However, it is quite unlikely that real-world vocabularies are fully semanticized or totally delexicalized; rather, they may be interconnected in varying ways that could prove characteristic of different stages (or types) of L2 collocation knowledge development.

What affects such varying interconnectivity is an issue we will return to later in the discussion when we consider how lexical knowledge and use interrelate with L2 collocation development. For the moment, however, let's consider how we can interpret the difference between the loadings for the conventional stereotypical collocation of standard use and the unconventional idiosyncratic collocation of the learner. One way is to focus on repeated use. Another concerns impermanent association. Repeated use is a principal means by which delexicalization is achieved by a community of users. Through repeated use, collocations become established. This is what characterizes native-like ability in that two items become strongly associated with each other, to the point of automatic retrieval and production. Repeated use and association then lead to semantic attrition in one component or delexicalisation. Early learner collocation production, in contrast, features impermanent association where two items may be temporarily linked for immediate communicative purpose, but do not become strongly associated with each other and permanently fixed as a retrievable

collocation. They are minimally repeated so that the bonds of collocation association within the L2 lexicon remain temporary at early stages of L2 collocation development. Because the bonds are temporary, it is difficult to speak of semantic loss in such learner collocations. In early stages of learning L2 collocations, collocation components do not undergo semantic attrition; rather, they retain their full lexicality, i.e., they are fully semanticized and stay independent of each other. Sinclair comments on this paradox with regard to native speaker intuition, but in fact his comments would seem to apply to L2 learners too:

> The 'core' meaning of a word—the one that first comes to mind of most people—will not normally be a delexical one. A likely hypothesis is that the 'core' meaning is the most frequent independent sense. (Sinclair, 1991, p.113)

Holding on to full semantic meaning may thus delay or even prevent L2 collocation development. Moreover, maintained lexicality is most likely to be located in the primary decontextualised sense of a particular lexical item—a point that we will discuss further in the next section.

Overall, we may argue that the collocation challenge for the L2 user is, in part, to surrender meaning—to give up the semantic certainty of individually acquired and temporarily associated items in order to develop L2 collocation ability. Kjellmer puts it like this:

> The learner...having few automated collocations, continually has to create structures that he can only hope will be acceptable...His building material is individual bricks rather than prefrabicated sections. (Kjellmer, 1991, p.124)

However, it is unlikely that the shift from lexically full items to delexicalised combinations may take place without there being some kind of change in the organisation of frequent lexical items within the L2 lexicon. This change would at least include the creation of more (and ultimately more) permanent associations and,

therefore, more automated collocations. Indeed, re-interpreting how the organisation of the L2 lexicon may change is one of the two dimensions that Meara (1996) proposes as important for understanding lexical competence:

> ...the size dimension becomes less important, and...the organisation dimensions might increase in importance as the size of the lexicon gets larger. (p.50)

What that organisational change might involve depends, as suggested above, partly on the links between frequent lexical items within the L2 lexicon. It is also governed by how infrequent lexical items are linked or not to frequent lexis. We will consider these aspects of L2 collocation knowledge and use next.

10.4.2 The combination of frequent and infrequent lexical items in L2 collocation use

I begin this part of the discussion by noting some problems in classifying frequent and infrequent vocabulary. These problems concern polysemous items and core decontextualised meaning. I then present an argument for seeing frequent and infrequent vocabulary in terms of their associative links with other lexical items. This may enable us to understand these associative links as contextualizing connections, or, in other words, as the potential basis for collocation. I conclude by exploring how frequent and infrequent lexical items may overlap or differ in the ways that they become linked and collocated.

Until now, we have taken a rather uncritical view of what we mean by frequent and infrequent vocabulary. This has been a necessary part of having some broad criteria for classifying individual responses in decontextualised productive tests of word association and collocation into two categories, so that we could identify differences between low proficiency and high proficiency learners. The broad division between frequent and infrequent was useful, and it worked from the premise that infrequent

vocabulary lies beyond approximately 2000 common word families of English. This does not seem an unreasonable position to take, given that the 2000 figure covers about 80% of the vocabulary in a typical text (Nation, 2001, p.14), although the coverage varies according to the type of text and may be as high as 90% for fiction (Nation & Kyongho, 1995, p.35). Frequent vocabulary, as we have used the term, is made up of two sets of vocabulary, K1 and K2 items. If we lexically profile most kinds of learner productions, the great majority of the frequent vocabulary will belong to K1 items, or the one thousand most frequent. In most cases, probably less than 10% of the frequent vocabulary will consist of K2 items, i.e., items from the second thousand most frequent word families of English. Nation and Kyongho (1995) put the figure quite conservatively at between 4-5% for K2 items. So, when we use the term frequent vocabulary both for texts and learner productions, we are really referring to about 1000 word families. This re-adjusted figure is important because it narrows down the dimension of size and suggests it is not simply how many words that learners know, but rather what they do with the very frequent words that they do know. A low proficient learner may know the 1000 word families in a different way from a more proficient learner. Knowing, in this sense, points to differences in the way learners' L2 mental lexicons are organized, so once the size dimension is adjusted downwards, the other dimension of organization assumes greater importance in the way that we understand the development of the L2 lexicon.

The second problem with our general understanding of frequency concerns polysemy. Polysemous items may skew our understanding of frequent K1 lexical items and obscure important aspects of lexical development. Bogaards and Laufer (2004) note:

> Words with several meanings, polysemes or homonyms, may appear higher up on frequency lists than monosemous words by virtue of the combined frequencies of their multiple meanings. Hence the content of these lists cannot be taken to be homogenous in terms of learner tasks. (p.x)

A relevant example from the General Service list is **address** (West, 1953, p.7). This has a K1 lexical profile (Cobb, 2003). It occurs 431 times in the GSL corpus, of which 42% of occurrences are covered by two nominal uses (**on a letter** and **speech**), and 49% by three verbal uses, namely **address a letter** at 12%, **address a person** at 16% and **address the King, a meeting** at 21%. Each individual semantic count suggests that particular senses of **address** are much less frequent than the aggregate polysemous total indicates. A learner is unlikely to learn all the different uses of the same lexical item at the same time. Let us assume that the nominal use of **on a letter** is the one that a learner is most likely to acquire first for **address**. (We discuss the precise reasons for this tendency in the next section when we consider the centrality of noun knowledge.) We will assume also that the learner does not simply learn frequent vocabulary through translation into Japanese and that there is some learning done in the medium of English. This points to some lexical items, if remembered by an L2 learner, developing at least one connection to another lexical item in the L2 lexicon. In the case of **address**, such a minimal connection is likely to be an associative extension, perhaps **letter**, or maybe a clang association such as **dress**. The association of **letter** with **address** can be called a decontextualised meaning in that it reflects our knowledge of the world, but does not mirror how we use **address** in context. *Collins Wordbanks Online* (HarperCollins, 2004) reveals that **letter** does not occur as one of the 100 most frequent collocates of **address**. Table 10.6 below shows the most frequent 20 collocates.

Table 10.6 **Top 20 collocates of address (in order of frequency)**

name, telephone, your, and, logo, number, send, addressee, dear, please, address, form, contact, supplied, office, change, write, issues, postcard, mail

The most frequent contextualised uses of **address** occur in a phrase such as **your name, address and telephone number**. What this implies for the learner is that learning the decontextualised, quasi-encyclopedic meaning of **address** takes them away from developing contextualized associative links with collocation potential. Formal learning leads to alienation from real-world use, in other words. If this is a common tendency in the acquisition of other frequent lexical items, then it is quite probable that the first stage of organization of the L2 lexicon may involve the development of many isolated strands between two or three lexical items at most, where concrete decontextualised senses predominate and lexicality is maintained.

A second stage of lexical development can be characterized as a partial re-organisation of the L2 lexicon. For this, at least two other processes need to come into play. In the first process, we assume that the learner starts to create further links between some already minimally interlinked lexical items. These then become multiply-connected to other lexical items, although most lexical items continue to be weakly connected. At the same time, a second process of de-activation and growth may be assumed to take place, where some already acquired frequent vocabulary becomes inactive or passive, and some new, infrequent vocabulary (i.e., outside the K1) joins the lexicon and forms active links with other active items. It is likely that some of these new infrequent lexical items may be linked to those few very active parts of the learner's L2 lexicon that have already started developing multiple links. In this scenario,

the multiply-linked lexical items, some of which are connected to both frequent and infrequent items, may start to lose their lexicality, or independent semanticization, and become progressively delexicalised. As delexicalisation takes hold, sporadic collocation development becomes possible.

However, two basic contrasts can be identified in the behaviour of frequent and infrequent lexical items. First, frequent lexical items have multiple senses and can collocate with a relatively large range of other lexical items, whereas infrequent lexical items have fewer senses and can collocate with a relatively small number of other lexical items. Second, frequent lexical items can collocate with both frequent and infrequent lexical items, but infrequent lexical items tend to collocate with frequent. Under ideal circumstances, a learner would be able to maximize links between frequent lexical items as well as connect infrequent items to them. Yet, we must set this idealization against the reality of ongoing growth and de-activation in what has been already acquired. So, there will be differing areas of interconnectivity within the L2 lexicon, with some parts sparsely, if at all, connected.

Summing up, we have taken a critical look at what the descriptors 'frequent vocabulary' and 'infrequent vocabulary' imply and argued that we need to adjust downwards what we mean by frequent lexical items. We also considered the issue of polysemy and claimed that, in early stages of learning, frequent lexical items are likely to be learnt on the basis of an encyclopedic understanding of language where some kind of decontextualised core meaning is prioritized over contextualized use. After outlining, albeit in rather sweeping terms, some major phases of L2 lexical development that impinge on collocation development, I suggested that infrequent lexical items probably become linked to frequent lexical items that have already started developing multiple

connections with other frequent lexical items. In short, it is the development of multiple active links that appears to foster delexicalisation, contextualized knowledge, and collocation potential. Finally, we looked at differences in the linking of frequent and infrequent lexical items within the L2 lexicon and suggested that lexical growth and lexical (re-) organization present the L2 learner with conflicting and, at times, irreconcilable demands. In the next part of the discussion, we narrow our focus towards the role of noun knowledge in L2 collocation development.

10.4.3 The centrality of noun knowledge in L2 collocation knowledge and implications for L2 collocation development

Results from different experiments pointed to the centrality of noun knowledge as the key anchor of collocation development. The data from the noun recognition experiment in Chapter 6 showed that noun knowledge was stronger than verb knowledge, and the experiment concerned with collocation production (in written data in Chapter 3) showed that nouns are the word form which L2 users most use for collocations. The final test of collocation production in Chapter 9 also showed that more proficient learners are likely to move beyond adjectival collocates and produce nominal-verbal collocates, or homonyms, at later stages of collocation development. To explore the question of noun centrality further, I will first refer to L1 acquisition research that points to the primacy of nouns in language development; I will then discuss relevant findings from L2 studies about noun learnability. This will let us draw out some interesting ideas about the development of multiple active links in the mental lexicon and understand the concept of varying interconnectivity in more specific terms.

Studies of L1 acquisition indicate that concrete nouns are the most frequent type of word in early child language (Dale, 1972; Gentner, 1982; Dromi, 1987). Animal

words such as **dog** and **cat**, objects in the immediate environment such as **milk**, **apple** and **car**, and proper nouns such as **Mummy** and **Daddy** are found to be common among the first words that children speak in different languages. They are also much more frequent than verbs. Gentner (1982) claims that concrete (and proper) nouns are the most accessible to the child learner, because they "have a particularly transparent semantic mapping to the perceptual-cognitive world" (p.328). That is, they offer the child a stable equivalence between object concepts and words. Verbs, on the other hand, do not map so easily for the child, and the child takes much longer to establish relational links between action concepts and particular verbs. Gentner's basic argument is that, in early first language acquisition, a fundamental difference can be observed between the rate of acquisition for concrete nouns and predicate words such as verbs and adjectives. She calls this position the Natural Partitions Hypothesis. As this hypothesis characterizes only the first stages of L1 acquisition, Gentner is careful to point out that as a child's language becomes more abstract and independent of the immediate perceptual world, the chances of conflating relational concepts become much stronger. Concrete nouns, Gentner concludes, are "an initial set of fixed hooks with which children can bootstrap themselves into a position to learn the less transparent aspects of language" (p. 329). Nouns, then, have a psycholinguistic reliability that other word classes may not.

Other evidence for this argument comes from studies of some types of aphasia where patients show greater use of nouns than verbs. According to Aitchison, this may be because "nouns are relatively free of syntactic restrictions" (Aitchison, 1994, p.102), suggesting that nouns are more easily accessible and easier to use than other word classes. Miller and Fellbaum (1991) further suggest that nouns have distinguishing

semantic features at different levels of generality which enables them to enter into richer lexical relationships than verbs or adjectives. There is, in short, quite strong evidence from L1 studies that nouns play a central role in language acquisition.

L2 studies have generally followed L1 studies and highlighted the learnability of nouns in terms of their imageability and availability. According to Ellis and Beaton (1995), imageability depends on direct, visual experience. Noun concepts can be easily related to "discrete perceptual experiences" (p.155), and the perceptual elements of a noun concept are densely interrelated to each other. Concrete noun concepts tend to package what is visually experienced into imageable, meaningful and memorable networks of interrelated features (Ellis, 1994b, p.252), so they are more meaningful than verbal concepts. The features of verbal concepts are sparser and less closely interlinked to each other. What is imageable is more available for access, too, and predictable. This is known as the "context availability" hypothesis. Hulstijn (1997) uses the context availability hypothesis to show how concreteness can help retrieval of words when prior contextual knowledge is lacking. He also invokes the imageability hypothesis to explain why concrete words "have the edge over abstract words in recognition, recall and lexical decision tasks" (p.213). These different L2 readings of the L1 research are very similar, but, in my view, downplay other findings relevant to our previous discussion of dimensions of L2 collocation knowledge and use—particularly the issue of the development of multiple active links within the L2 mental lexicon.

Relevant research here includes the work of Nelson and Shreiber (1992) who examined relationships between word concreteness and word structure. By word structure, they refer to "the number, strength and connectivity of a word's associates" (p.237)—precisely the features that we had discussed earlier as important for

understanding conditions that may help or hinder L2 collocation development. Nelson and Shreiber report on three experiments in which they manipulated concreteness and set size to see what effects these variables had. On the one hand, they found that concrete words were more memorable than abstract words—which is very much in line with what we have just summarized from L1 and L2 studies about the centrality of noun knowledge. However, Nelson and Schreiber also found that "words with smaller associative sets were more likely to be remembered than words with larger sets" (p.249). In other words, it is the compact organisation of the associative set rather than concreteness or abstractness that influences memory: "...concreteness effects cannot be attributed to differences in structure theoretically underlying concrete and abstract words" (p.256). It seems highly plausible that conceptual links (or small sets of associative links) are more readily available for nouns than they are for verbs or adjectives.

This has two important implications for the L2 collocation development. First, we may identify the centrality of noun knowledge as a paradigmatic phenomenon. Nouns, whether concrete or abstract, embody basic conceptual knowledge of the world. These basic concepts tend to be paradigmatically organized in relation to each other, and for these concepts to be differentiated from each other, sets of associative links need to be activated. For example, if we took the concept **buildings**, some particular features of the concept **house** would need to be activated in order to distinguish it conceptually from, say, **supermarket** or **school**. Many of these distinguishing conceptual links will be syntagmatically encoded, i.e., bound up with the way the particular noun is collocated (for example, **buy/sell/let/rent a house, detached ~, four-bed-roomed ~; be off/go to school, primary ~; go to the supermarket, local ~,**

~ checkout). It is also probable that many of the links will be adjectival, nominal or verbal, for it is lexical items in these word classes that realize the majority of collocations. In this sense, the adjectival, nominal and verbal collocation of nouns is a process of restrictively modifying and textualizing conceptual knowledge and ideological views of the world.

The second implication concerns our general understanding (at this point in the discussion) of how the L2 mental lexicon may be organized at different stages of development. L2 collocation knowledge seems to be related to how compactly interconnected different lexical items are. Let's recall here the differences that we found in Chapter 8 between Emi, Kensuke, and Shoichi in the way that they could repeat or not the same collocation links at different interventions. Emi, the most proficient of the three, was found to repeat collocation links much more than Kensuke and Shoichi, both of whom tended to activate different collocation links more often. Emi, in this sense, seems to have worked from smaller compact sets of associative links that allowed her to recall appropriate collocations more successfully. Kensuke and Shoichi's associative sets were much more diffuse, although, in Kensuke's case, he activated a greater number of different collocates and, at the same time, produced more inappropriate collocates. His L2 lexicon was, it appears, diffusely organised and loosely linked, whereas Emi's was more tightly organized. The ability to reproduce small sets of collocation links demonstrates L2 collocation ability and is also an indication of conceptual differentiation (see 10.5.1 below on the bilingual lexicon and conceptual shift for further discussion of this).

Overall, there are reasonable grounds for believing that nouns are more likely than verbs or adjectives to enter into durable associative links with other lexical items,

despite the individual differences that we have just noted. Nouns are conceptually richer and more compact than verbs or adjectives. As far as L2 lexical development in general is concerned, it seems quite likely that learners may create associative links between nouns and other lexical items. This can, in the longer term, create favourable conditions for the production of L2 collocations.

10.4.4 Summing up the discussion on exploring dimensions of collocation knowledge and use

The discussion so far has led us to develop our understanding of dimensions of L2 collocation knowledge and use in the following stages. As the L2 mental lexicon develops, lexical items start to become multiply linked. At first, lexicality is maintained and items are linked in decontextualised encyclopedic senses. When learners start to create their own idiosyncratic associative links, frequent lexical items become more multiply-linked in the L2, and this provides the basis for initial delexicalisation of parts of the lexicon. The restructuring of the L2 lexicon through the spread of delexicalization is neither linear nor uniform and occurs sporadically where frequent lexical items are repeatedly re-activated. These items are most likely to be nouns, some of which will become anchors for new infrequent vocabulary in the L2. The critical factor in priming the further development of L2 collocation knowledge and use is the balanced re-activation through repeated use in the L2 of both frequent and infrequent vocabulary. Exactly how that priming of L2 collocation development might be achieved is a central focus in the next stage of our discussion.

10.5 Linking lexical knowledge and use with collocation development

The priming of L2 collocation development is a key part of the continuing discussion: what can provide conditions for this, and why? Here I will consider the priming role of

associative knowledge and paraphrase in L2 collocation development, and how such lexical priming affects the interconnectivity of collocated items in the L2 mental lexicon. I will also examine the role of context, so that we can begin to examine L2 collocation development within local situations of L2 use, and not just in psycholinguistic terms. However, before we tackle those more contextual issues, we need to re-consider the nature of the bilingual lexicon and differences between native-like and L2 collocation ability.

10.5.1 The bilingual lexicon: network development and conceptual shift

Much of our discussion has focused on trying to understand lexical and collocation development as network building (Read, 2004). As the term implies, network building focuses attention on the development of the L2 lexicon structure as a whole rather than on "a collection of individual words and meanings" (Meara, 1992, p.69). Read characterizes the development of a lexical network in the following way:

> ...as a learner's vocabulary size increases, newly acquired words need to be accommodated within a network of already known words, and some restructuring of the network may be needed as a result. (Read, 2004, p.219)

For Read the fundamental feature of the network model is the creation and accommodation of connections between different words, and his description pays particular regard to how freshly learnt vocabulary is dealt with. Our discussion has similarly suggested that the restructuring of a lexical network is not just a question of what happens with what is added to it, but also what occurs with vocabulary that is already part of the network. In framing lexical network development in terms of repeated activation, spreading delexicalisation and restructuring of the L2 mental lexicon, one of our underlying assumptions is that development implies ever greater achievement of native-like collocation ability in the L2. However, this idealization of

native-speakerness as the end-goal is problematic in a number of ways. The most obvious problem is the premise that people with two languages should be compared to people with one, as if it is the monolingual native speaker standard that bilinguals should be judged by. A second related problem is how we can judge L2 collocation ability if we do not use a native-speaker standard to assess collocation appropriacy: just exactly what do we want to mean by L2 collocation ability?

Clearly, one of the most basic lexical differences between a monolingual native speaker and a bilingual (in the broad sense of someone who tries to learn and use a second language) is that the monolingual has at their disposal a single language for mediating their conceptual world and use of the lexicon. A monolingual has direct access between conceptual and lexical links so that they have more or less complete control over their ability to produce appropriate language to express what they want to mean. Despite this rather ideal view of how functionally proficient monolingual native speakers are, the important point is that a NS can map lexical items directly and consistently to intended meaning: one result, among many, is that appropriate collocations can be automatically produced. A monolingual user of English may effortlessly produce the collocation **take medicine,** whereas a bilingual user may, for example, have access to the English **medicine** in their minds but hesitate to collocate it—in the case of a L1 Japanese-L2 English user, this might involve **drink medicine** (**kusuri o nomu**) or **put medicine** (**kusuri o tsukeru**) if they draw on their L1 conceptual and lexical knowledge. For a Japanese person, the action of putting a medicine in their mouth is close to the concept of drinking and swallowing something, so until the L1 Japanese-L2 English user modifies their conceptual view of the world, it is likely that the production of the collocation will continue to be subject to L1

conceptual mediation. Thus, the collocation will be realised as an expression of that conceptual view (i.e., **drink medicine**). A shift in conceptual mediation from the L1 to L1+L2 or L2 alone appears to be necessary before the native-speaker collocation **take medicine** can be realised. When that conceptual shift has taken place, different L2 lexical links can be activated, so that the production of an appropriate L2 collocation is possible. I am invoking, in other words, three levels of representation (conceptual, lexicon, and activated lexical links) in order to articulate a model of bilingual competence relevant to L2 collocation development.

Bilingual studies support such a view, notably Ringbom (1987), Kroll (1993), and Kroll and De Groot (1997). Kroll and De Groot argue that bilinguals become less dependent on direct L1 conceptual mediation as they become more fluent in the L2. In the early stages of learning an L2, the learner links L2 lexical items "primarily by links with the first language" (Kroll & De Groot, 1997, p.179) and it is only later that "direct conceptual links are also acquired" (Kroll & De Groot, 1997, p.179). This does not mean that L1 links are erased; rather, they remain as possible bonds, apparently at a subliminal level of conscious control. Kroll and De Groot (1997, p.190) put forward the distributed lexical/conceptual feature model to illustrate how words in two languages may have common lexical and conceptual features, but be variably activated according to the stage of development, context and task. The model has three main components: lexical features, L1 and L2 lemmas, and conceptual features. If we re-label the lemmas component as *lexicon*, and the lexical features component as *activated lexical links*, we can align the model to show how conceptual, lexical and collocation knowledge may variably interact.

An example representation is shown in Figure 10.1 below for an early stage of L2 development. In Figure 10.1, the black shading indicates strongly activated links, and the black lines show how the strength of the activation between components. Similarly, the grey shading and white circles represent less activated links in both the conceptual and lexical components, and the weaker dotted lines indicate very weakly activated connections between the different components.

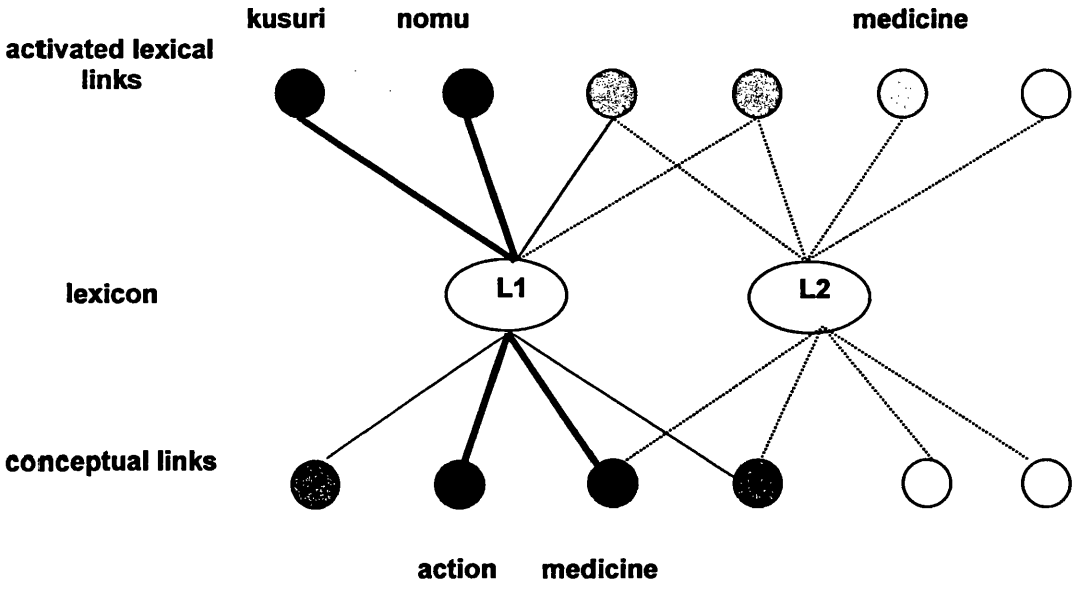

Figure 10.1 L1 conceptual mediation: L1-linked collocation

Figure 10.1 illustrates L1 conceptual mediation and L1 activation of lexical links: the collocation produced is in Japanese. Although **medicine** is weakly activated as a lexical link in English, the noun is not collocationally linked, and L2 collocation production does not occur.

Figure 10.2 below, on the other hand, shows a different interplay between L1 conceptual mediation and L2 activation of lexical links. Here, there is mixed L1/L2 mediation where the L1 concept is still dominant in its activation of both L1 and L2

lexical links. Under these circumstances, a transparent collocation is produced (**drink medicine**).

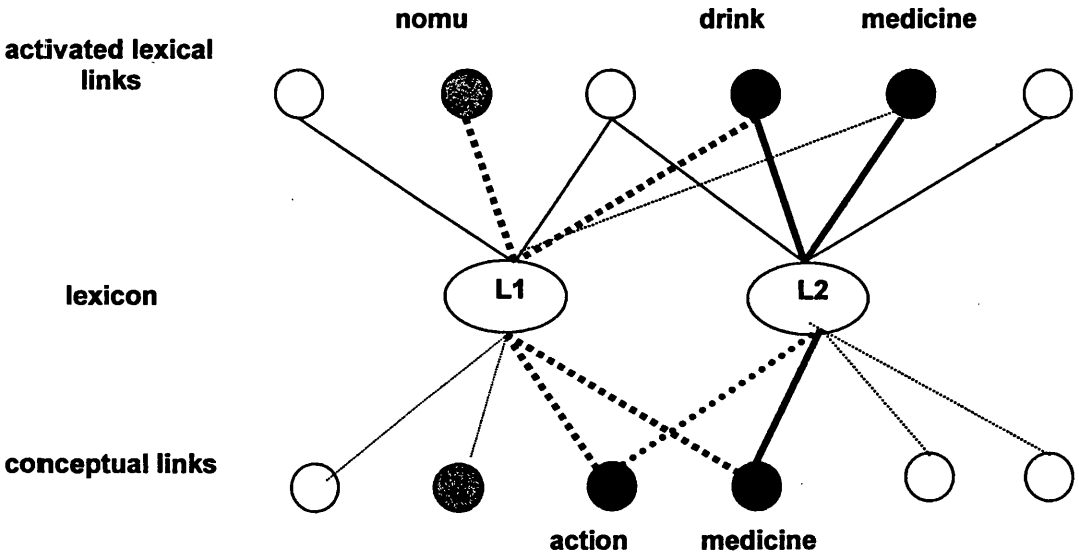

Figure 10.2 **Mixed L1/L2 conceptual mediation: mixed L1/L2 collocation**

In contrast, Figure 10.3 further below shows a shift in conceptual mediation from mixed L1/L2 to the L2. This then allows for L2 lexical links to be activated, so that the production of an appropriate L2 collocation is possible.

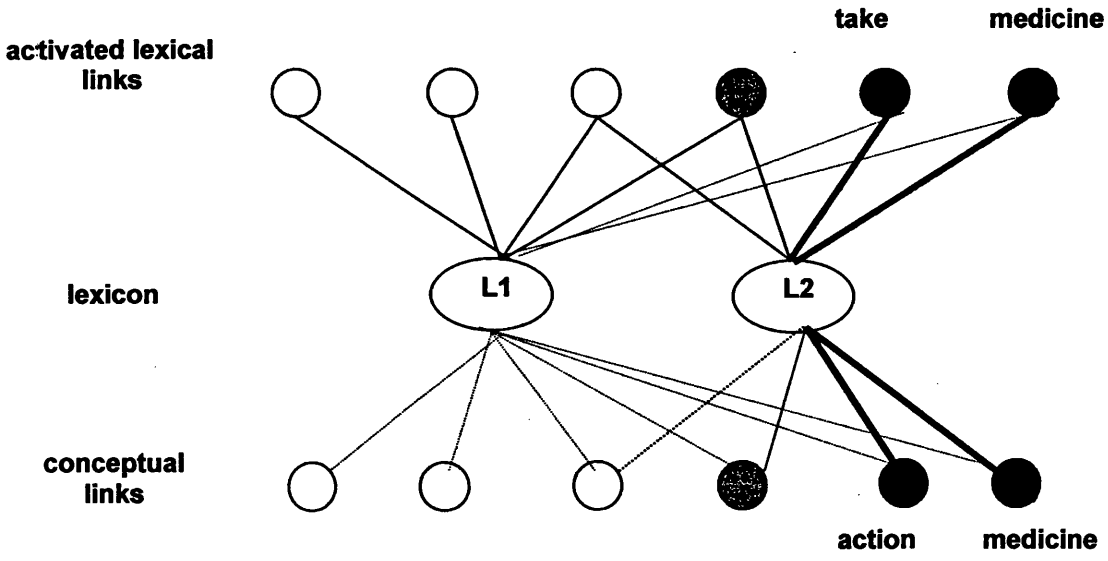

Figure 10.3 **Conceptual knowledge mediated by the L2: L2-linked collocation**

In Figure 10.3, we have a representation of what we earlier referred to as conceptual shift through restrictive modification. Through L2 conceptual mediation, a different view of the world is textualized, where basic paradigmatic knowledge is modified through the activation and use of syntagmatic lexical links.

Although the development of L2 collocation knowledge involves conceptual shifts, as well as repeated activation of lexical links, for the bilingual language learner/user, we need to be careful to qualify this interpretation. Earlier, I claimed that the restructuring of the L2 lexicon through the expansion of delexicalization is neither linear nor uniform. If conceptual shift from the L1 to L2 also underlies L2 collocation development, then this must also be uneven and relative to particular learners, contexts and uses to which they put the L2. It is perhaps then more suitable to see the conceptual dimension that underlies the development of L2 collocation knowledge as a range of potential discrete conceptual shifts that may be contextually triggered. Most learners in EFL contexts (such as Japanese university students) have highly limited exposure to English as a second language, where they may use English for less than two hours a week. Contextual triggering in such low exposure environments will often be scarce, and collocation development weak, unless the individual becomes particularly language-hungry and spends some considerable amount of time learning and using English outside the classroom. Thus, we can expect great variation in individual productive collocation knowledge as the results from the COLLPROD measure showed. Additionally, we cannot assume that mere exposure and use are sufficient conditions for triggering conceptual shift and productive collocation development. As Jiang notes (2004), if an L2 lexical item is linked to a L1 concept, it is quite possible that repeated activation will strengthen the L1-L2 connection rather than lead to a shift to L2

mediation: "Restructuring will successfully take place only when the conflict between the transferred semantic understanding and the understanding obliged by the context is powerful enough to override the transferred meaning..." (Jiang, 2004, p.105). There are then a host of other factors to consider in understanding how individual L2 learners manage the shift and create spaces for collocation development, and we will shortly come to the more important of these in the following stages of our discussion.

The second point to note is that the examples presented in Figures 10.1-10.3 are of a single collocation, but a conceptual shift is likely to involve a somewhat larger mini-network of conceptual features and small sets of activated lexical items. The example figures were presented to illustrate a more complicated process of change and should not be read too literally as always referring to the acquisition of particular collocations in isolation. The final comment that I would like to make is that the type of conceptual shift shown in Figure 10.3 (i.e., conceptual knowledge mediated by the L2 and L2-linked collocation) is close to a view of L2 collocation ability that involves intermediate plus English proficiency—that is, learners who are able to draw on L2 resources independently of the L1 for expressing themselves. It is tempting, then, to see L2 collocation ability as an advanced form of lexical knowledge. However, if we do this, we risk losing sight of other aspects of L2 collocation development where the L1 and L2 mediate and interact. This brings us to the question of how to describe more precisely differences between native-like and L2 collocation ability, which we turn to next.

10.5.2 Differences between native-like and L2 collocation ability

The results for V+N collocation recognition in Chapter 6 and for collocate production with the 30-item measure COLLPROD in Chapter 8 demonstrated that learners from the

low-intermediate level upwards have L2 collocation ability. We found that recognition of V+N collocations was weaker than recognition of individual verbs and nouns, which suggested that learners store individual lexical items rather than their combinations. We furthermore established that L2 collocation recognition knowledge is not restricted to advanced learners. At the same time, the results showed that collocation recognition knowledge did not significantly correlate with general English proficiency in the mid-intermediate range. The significant differences that the test was able to measure were between a high-intermediate group and a low-intermediate group. On the other hand, the results from the test of collocation production let us characterize increased productive knowledge in a number of ways—the greater use of frequent and infrequent collocation links as mentioned earlier in this chapter; significant diversification in word classes for collocates, and significant correlations to overall English proficiency. All these features could be taken as signs of increasingly native-like collocation performance, even if the evidence from the recognition test pointed to decidedly non-native-like ability, i.e., strong analytic knowledge of individual lexical items. The results for L2 collocation recognition and production present us with something of a paradox in trying to define L2 collocation ability as becoming more native-like because the picture also includes signs of markedly non-native-like ability. The paradox lies in the fact that the L2 learner, "however accurate in grammar and knowledgeable at the level of words" (Wray, 2002, p.210), keeps lexical items separate and "the adult learner's collocations are to be seen as separate items which *become paired*" (p.211). In contrast, as Pawley and Syder (1983) argue, a native speaker retrieves stored familiar collocations as "wholes or automatic chains from the long term memory" (Pawley & Syder, 1983, p.192).

In face of this paradox, we may initially claim that the intermediate level of proficiency is where major changes in L2 collocation recognition ability need to occur if the second language learner is to make it through to more advanced (and native-like) proficiency. Indeed, in the light of our previous discussion of network development and conceptual shift, it is at the intermediate level that learners may start to break through to independent collocation development in the second language. L2 mediation, independent of the L1, is the key challenge for intermediate learners. Within this proficiency range, L2 collocation recognition should be stronger than production, with learners able to identify decontextualized collocations better than they are able to produce them. Collocation recognition ability does not correlate with overall proficiency because it focuses on isolated aspects of L2 lexical knowledge that can be L1 mediated. Part of the evidence here is that the 15 most recognized V+N collocations were almost totally lexically congruent with their Japanese equivalents. This also indicates that L2 collocation transparency is an important part of initial L2 collocation development—a point that I will expand on further below.

When learners are faced with producing L2 collocations, L1 mediation will be of less help. Because L2 learners tend to store lexical items separately, it is difficult to recall and produce native-like collocation choices. Part of the problem is naturally that collocation use involves reconstructing the arbitrary combinations that have become conventionalized in the target language. Many of these are not lexically congruent between languages, so the learner must re-create L2 collocations from individual lexical items (cf. Wray, 2002, p.209, on how the adult language learner deals with encountering, storing, and later reproducing the collocation **major catastrophe**). Such productive knowledge of the language involves drawing on a much wider range of lexical resources,

and learners with greater overall proficiency in English will show greater appropriacy. That is why productive L2 collocation knowledge is a better indicator of general language ability than L2 collocation recognition knowledge.

A different way of understanding the contradictory nature of the development of L2 collocation ability is to see it as involving identity conflict for the second language learner/user. In the case study in Chapter 9, Kensuke was the one learner whose collocation knowledge did not noticeably improve over the course of the interview period. Part of the reason for this was that Kensuke directed his vocabulary learning largely towards translating into Japanese new and difficult English words that he met. He continued to follow the way of learning English that he had been required to adopt in high school in preparation for the English part of Japanese university entrance exams. Students prepare particularly intensively for these entrance exams in their last two years at high school through grammar-translation and the rote learning of ever more infrequent English vocabulary lists. They are almost exclusively required to be able to supply the Japanese translation for each word they learn, so it is not surprising that they can recognize individual items quite well, but have a quite limited sense of how frequent and infrequent items are meaningfully combined with each other. Kensuke was invested in maintaining the identity of an English learner used to the formal external evaluation of his English ability from the L2 into the L1. He was not willing to make a different choice for himself and would not change his approach of L1 control to language learning. He deemed his actual knowledge sufficient and was resistant to developing the potential of his L2 ability.

What happens, though, if a learner changes their understanding of language learning away from "school knowledge" (Barnes, 1976)—i.e., what is required to be

learned in formal classroom settings and is evaluated through institutional testing—towards "action knowledge" (Barnes, 1976) where they appropriate the language for their own particular communicative purposes? What occurs if an individual, for example, becomes a more self-directed user of the second language, goes to an English-speaking environment for a short period, uses English actively, then returns to Japan and re-enters a low-exposure English environment? In the case study, in contrast to Kensuke, Emi did just this and showed a gathering sense of identity as a reflective language user—as somebody who was not content with simply memorizing language, but who wanted to learn and use English for her own purposes. She became highly conscious of different ways in which she could organize her L2 collocation development successfully for herself. Emi had a very strong interest in continuing to develop her English in the future and hoped to study human rights at the graduate level. It is clear that Emi had a growing sense of a future professional identity for herself in which the fluent use of English would be extremely important. Equally, she transformed her sense of herself in the present, both inside the classroom and outside, so that learning and using English became dynamically interconnected. She learnt English by doing human rights research and rehearsing to herself, as much as she used English to explain her research to others: Emi became increasingly focused on realizing her potential L2 collocation knowledge.

Yet, students like Emi are not typical of the majority, and many get stuck in the middle, so to speak, as they shift back and forth between the conflicting identities of learner and user and confront breaking away from school knowledge habits that the education system has demanded of them. This is one way in which to re-interpret Shoichi's fluctuations. Certainly, he struggled with finding ways to direct and control

his L2 collocation development, and without doubt part of this was affected by his initial focus on studiously checking new and difficult words. Later, he was able to discover his own way, where a noticeable change was his focus on learning and using useful and important vocabulary and collocations for explaining his ideas to others. Shoichi struggled to switch identities, but, at the end of the interview period, had moved from formal L2 learner to active L2 learner-user.

Emi shared a similar drive to explain clearly to others. Both Shoichi and Emi were concerned with learning and using transparent collocations to share and present their human rights research. These social and affective dimensions of their involvement—notably missing in Kensuke's disengagement from investing himself in action knowledge—were critical in Emi and Shoichi's development, as Little reminds us: "It is...important to emphasize that all learning proceeds via interaction, so that the freedoms by which we recognize learner autonomy are always constrained by the learner's dependence on the support and cooperation of others" (Little, 1997, p.230). The peer-oriented motivation to use transparent collocations—founded on a sense of communicative interdependence with others—takes us now to the next theme in the discussion, the priming role of associative knowledge and paraphrase for L2 collocation development.

10.5.3 The priming role of associative knowledge and paraphrase for L2 collocation development

One of the major insights that we gained from qualitative analysis of Emi and Shoichi's constructs of L2 vocabulary development was that they took two important lexical actions to sustain their L2 collocation development. The first was that they used paraphrase to explain clearly their ideas. Paraphrasing involved them in interconnecting

frequent and infrequent vocabulary so that they re-activated and connected what they already knew to present to others what was new and important to them. The second was that the decision to paraphrase required them to use concise and transparent collocations so that they could comfortably explain their ideas to others and be confident that others would understand what they wanted to say. In other words, paraphrasing was the bridging mechanism between their *actual* individual, associative lexical knowledge and the *potential* development of their socialized collocation knowledge. Without the will to paraphrase content that was important to them, they would also have flat-lined in their collocation development as Kensuke did. By addressing the problem of explaining their ideas to others, they worked within their zones of proximal development (Vygotsky, 1935) at the edges of their productive lexical knowledge. Vyogtsky defines the zone of proximal development or ZPD in the following way:

> "It is the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers." (Vygotsky, 1935, p.86)

The critical point is that internal development processes can only be activated through interaction with others in the social world: participation is necessary in order for individual development to take place.

Yet, with participation come constraints, and, as far as L2 collocation development is concerned within the particular local EFL context of a content-based research and discussion course, the major constraint is the need to be understood by others. This in turn motivates the choice to use transparent L2 collocations based on lexical items that L2 learners are comfortable with (i.e., are established in their lexicons and can be retrieved for combination). Thus, Shoichi rejected using **pact** (a new

infrequent word) and chose instead to collocate the more familiar item **arrangement** with **suicide** to arrive at **suicide arrangement** for presenting his research. Interestingly, **suicide pact** is given in the Oxford Collocations Dictionary (OUP, 2002), but **suicide arrangement** is not. The perceived need to be collocationally transparent, in effect, restricts the development of native-like L2 collocation ability. To put it another way, L2 collocation ability is socially constrained by the learner's here-and-now socio-cultural context of interdependence with their non-native peer group: L2 users may have other more important goals than the achievement of complete, idiomatic native collocation ability. Rather than wanting to achieve native-speakerness, they are more invested in discussing the content of their learning with each other and co-constructing their knowledge of the world through the L2.

This has further ramifications for how we understand the restructuring of the L2 lexicon that paraphrasing can lead to. Individuals such as Shoichi and Emi link into their L2 networks those lexical items which they decide are useful for explaining their ideas. At the same time, they connect into their L1 networks those lexical items which they decide are unimportant for their purposes. To do this, they judge for themselves the relative frequency and usefulness of the vocabulary that they attend to. Emi, for example, decided to make no further connections in English for **strap**, **look-out**, and **amalgamation**, and in her notes simply wrote down the Japanese equivalents. On the other hand, both Shoichi and Emi reported difficulty in collocating highly frequent everyday lexical items (such as **advice**) because they did not have the opportunity to use such items, even if they knew they were highly frequent. Further evidence for this claim comes from the Rasch analysis of COLLPROD responses in Chapter 8 where **body** proved to be the second most difficult item to collocate.

All this suggests that, within a low-exposure EFL environment, the most likely lexical items for restructuring in the L2 lexicon are neither highly frequent everyday vocabulary nor highly infrequent vocabulary. It is rather those items that are 'somewhere in the middle' where L2 collocation development is likely to occur. 'Somewhere-in-the-middle' includes lexical items which are repeatedly encountered in English without being already conceptually over-established in the L1 (as they would be available for L2 mediation and lexical activation), but excludes lexical items which are so frequent so as to be part of the young adult learner's everyday personal world (as they would be strongly connected into the L1 and resistant to L2 mediation). Within the particular context of a taught-in-English academic-literacy programme in a faculty of law, these 'somewhere-in-the-middle' lexical items are basically the currency of an engagement with general academic issues and social, political and legal problems. Learners will have a strong chance of repeatedly encountering such vocabulary on a particular issue, as well as need to re-use, re-activate and re-combine selected lexical items in order to explain their ideas to each other.

10.5.4 Summing up linking lexical knowledge and use with collocation development

In this part of the discussion, I have explored both psycholinguistic and contextualized aspects of how lexical knowledge and use may inform L2 collocation development. Psycholinguistically, changes in conceptual mediation help or hinder L2 collocation development. The L2 user of English may exploit L1 conceptual mediation to activate lexical links in the L1, use mixed L1-L2 conceptual mediation, or draw on L2 mediation of conceptual knowledge. Although it would be appealing to see such conceptual mediation as sequential linear stages of overall collocation development, it is unlikely

that learners follow such neat stages of development. Rather, all three types of conceptual mediation are available from intermediate proficiency onwards for individuals to make use of. The specific types of mediation used by individual learners will vary. Intermediate proficiency is the band in which the learner faces the challenge of breaking through into increasing L2 conceptual mediation and L2 (re-)activation of lexical links, and individual variation will be strong as the results with COLLPROD showed.

L2 conceptual mediation and L2 (re-)activation of lexical links are a basic psycholinguistic condition for developing L2 collocation knowledge contextually. Through contextual use, collocations can become established. Here, two lexical processes are central. First, paraphrase provides a way for learners to restructure individual lexical knowledge socially. Such socialized lexical use will be governed by the need to be understood by peers and will therefore foster the second process, namely the production of transparent collocations. Transparent collocations are prioritized because they will be understood within the local context, and learners will be able to develop their knowledge of the world by using them. However, many transparent collocations may not be native-like.

Such peer-mediated interaction, under ideal conditions, takes the learner to the edge of their ZPD. It also involves the learner making increasingly informed choices about the type of vocabulary that they wish to use. The selection of useful and important vocabulary will tend to exclude highly infrequent lexical items, which may be simply linked into the L1 and rejected for collocation purposes. It will also tend to exclude highly frequent lexical items as these will be already tightly linked into the L1 through previous L1 conceptual mediation.

In sum, the zone of L2 collocation development covers vocabulary that is already partially established in the L2 and that can be re-activated as part of a general engagement with academic literacy through English. As the learner becomes engaged with their collocation development, they may move from a preoccupation with L1 translation and formal evaluation of school knowledge towards combining school and action knowledge for their own purposes. The stronger this shift is, the more likely it is that the learner will start to see learning language and using language as closely intertwined. The learner is therefore likely to go through some kind of identity change or conflict as part of their L2 collocation development.

Having summarized the psycholinguistic and contextual connections between lexical knowledge and use and L2 collocation development, I will now explore how the different strands of the discussion can be brought together in an emerging model of L2 collocation development.

10.6 Proposing a model of L2 collocation development

So far in this chapter, we have explored several key aspects of L2 collocation development. These include:

- moving from lexicality to delexicalization through repeated use and association
- variably interconnecting frequent and infrequent lexis
- adapting nominal concepts in paradigmatic lexical knowledge through syntagmatic modification
- developing variable L1-L2 relationships between conceptual links, the lexicon and activated lexical items
- transforming individual associative knowledge through paraphrase into socialized and conventionalized transparent collocation knowledge
- increasing sophistication in lexical decisions about the type of vocabulary to be collocated

- switching identities from formal learner to communicative user-learner.

In this final part of the discussion, I would now like to consider how we can relate these different aspects to each other.

10.6.1 An emerging model of L2 collocation development

Our discussion of issues affecting L2 collocation development has been both psycholinguistic and contextual. Because collocation knowledge is realized through use of language and cannot be restricted to a purely psycholinguistic interpretation, the contextual dimension of L2 collocation development needed to be taken account of. This led to a necessarily complicated view of what learners do on the inside and outside as they nurture their L2 collocation ability.

The emerging model that I will propose therefore has two fundamental dimensions, the psycholinguistic and the contextual. Included in the psycholinguistic dimension are those aspects of L2 collocation development that are important in the internal structuring of the L2 lexicon: conceptual mediation (i.e., L1, L1-L2 and L2), types of lexical organization (i.e., lexicality, sporadic delexicalization and increasing delexicalization), effects on the L2 lexicon from restructuring (i.e., L2-L1 linkage, loosening of L1 lexical links for L2 lexical items, re-activation and combination of lexical items) and the centrality of nominal concepts in initial syntagmatic modification of concepts and L1-L2 conceptual differentiation. These internal components of the model may combine in varying ways to block or enable L2 collocation development.

On this basis, such development has these probable observable characteristics at the intermediate level of proficiency: L1-linked collocations, L2-L1 mixed collocations, and transparent L2 collocations. The emerging model of L2 collocation development is presented in Figure 10.4 on the next page.

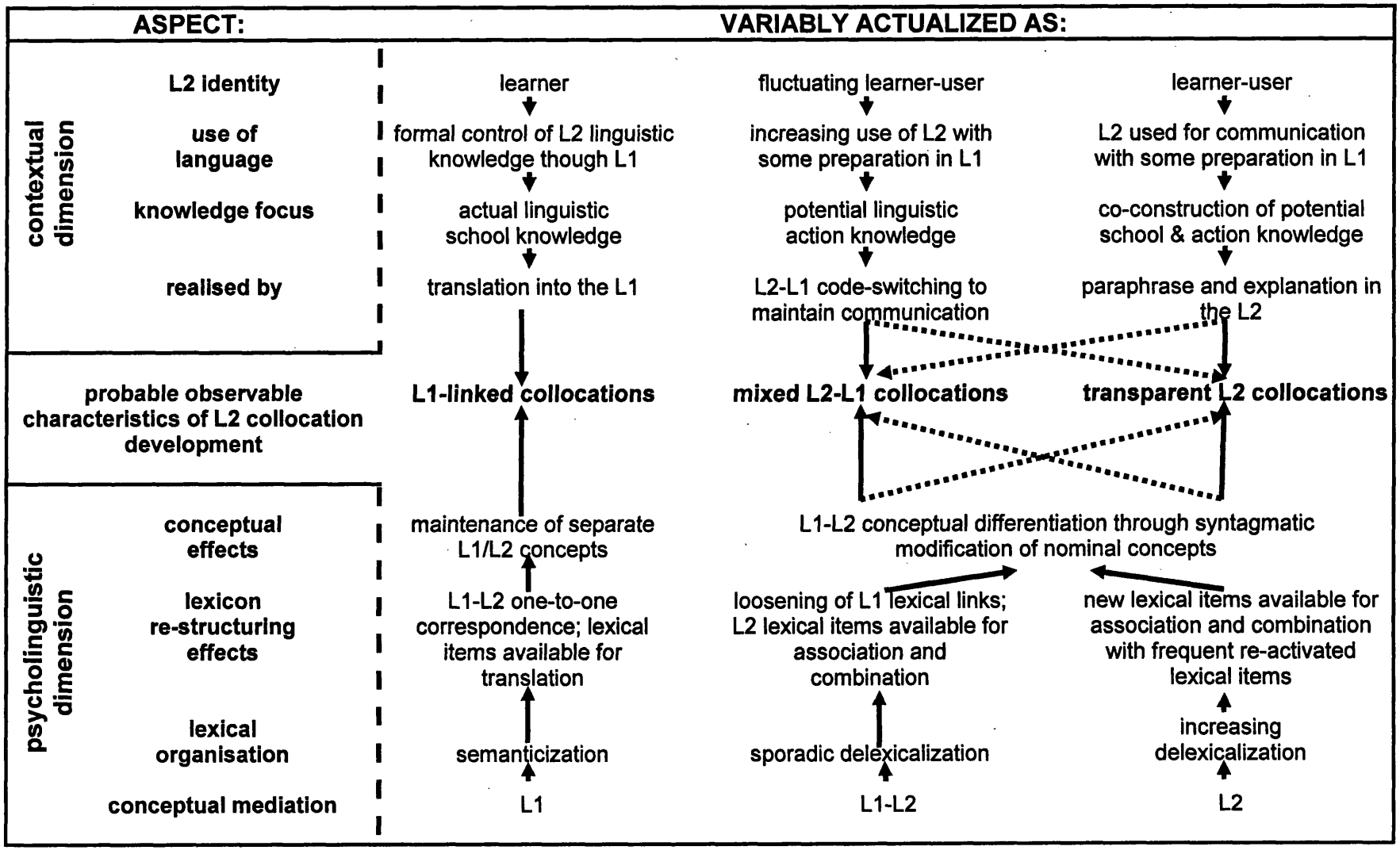

Figure 10.4 **An emerging model of variable L2 collocation development**

In the model, the observable linguistic behaviours mentioned above are also contextually motivated. In the contextual dimension of the model are included critical factors in the external structuring of L2 collocation production: L2 identity (i.e., learner, fluctuating learner-user, and user), language use (i.e., L1-based or L2-oriented), knowledge focus (school, action, or both), and the realisation of such knowledge (translation into the L1, L2 communication, or paraphrase and explanation in the L2). Again, these components may variably interact.

10.7 Conclusion

In this chapter, I have recapped the insights from the experimental work and explored different issues that the research into L2 collocation recognition, production and process has led to. The discussion has pivoted on first trying to identify plausible connections between psycholinguistic and contextual dimensions of L2 collocation knowledge, and second showing how these connections may be understood in an emerging model of variable L2 collocation development. The model is complex and clearly suggests that the development of L2 collocation knowledge is a multi-faceted phenomenon that cannot be easily reduced to a single area of interpretation.

In explicitly re-connecting with the complex character of collocation, the discussion has concluded by echoing the tensions that we observed in L1 collocation research. There, we focused on three major interpretations—the phraseological, the typological and the textual—and realized that no single interpretation could fully describe or account for collocation. Here, we started by focussing on the psycholinguistic, but were drawn to the contextual as we started to consider the situated development of L2 collocation knowledge.

In the end, neither dimension proves to be sufficient in itself for accounting for L2 collocation development, although each may be indirectly informed by the other.

In sum, the discussion of the research in this thesis has raised some complicated questions about vocabulary which do not fit easily with current models of L2 lexical acquisition and which will need to be the subject of further research. In the final chapter, I will very briefly outline some areas that may warrant further investigation.

Chapter 11

Conclusion

My very starting concern in doing the experimental work for this thesis was to find appropriate ways to explore the development of L2 collocation knowledge. I did not want to limit my work to developing over time a single measure of L2 collocation knowledge. Rather, I was interested in examining from different angles how L2 learners recognize, produce and learn collocations. As we saw in Chapter 2, quite a lot of previous sentence-level and corpus-based research had been concerned with learners' accurate production of collocations. Yet, very little work had been done experimentally to examine how well learners recognize collocations and how they may go about learning collocations for themselves. In this chapter, I would like to cast a brief but critical look at these areas of production, recognition and process, so that I may suggest some areas for future investigation

At the beginning of my research, I felt the tremendous pull of a corpus-based textual approach to investigating second language knowledge and use of collocations. What other way could there be to investigate collocation production than by using a corpus? Yet, my first attempt to understand L2 productive collocation knowledge showed that it would be very difficult indeed to move beyond a few examples per individual of their actual use of collocations, unless I took a different approach. The investigation in Chapter 3 of the collocation environments of lexical verbs was both a start and a turning point. It highlighted the tension in L2 collocation use between frequent and infrequent lexical items, but it did not provide a satisfactory basis for believing that a corpus approach was the route to follow for understanding this further.

At the other end of the exploration was the simple 30-item COLLPROD test, where learners had to supply three collocates for each stimulus noun. This measure may be closely related to word association studies in its design, but it stands in stark opposition to nearly all previous research into L2 collocation knowledge. COLLPROD proved to have a very high item reliability, but showed greater variation in its person ability. In a sense, then, COLLROD shows that it is possible to design a simple decontextualized measure of productive L2 collocation knowledge, if we accept that individuals will vary in their ability to respond to the test. There is nothing surprising about this, given that variability emerged as a major factor in the longitudinal case-study of the three learners' word association and collocation performance in Chapter 7. It remains to be seen whether the variability observed with COLLPROD would persist with a larger population.

One important area then is in the future for COLLPROD or a similar instrument to be tried with larger numbers of learners. As the instrument was designed by comparing the responses of highly proficient Japanese users of English with those of English native speakers, COLLPROD may well reveal further limitations if used with different L1 groups. A second area of attention would be to develop a computerized version of COLLPROD that could automatically process learner responses. The paper and pencil version of COLLPROD was highly time-consuming to work with, and there is no reason why the databanks of appropriate collocate responses could not be arranged as part of a computerized version.

One of the interesting aspects of COLLPROD was the diversification in word class categories that more proficient learners showed in their collocates. They tended to produce ambiguous responses in the sense that some of their answers could be

interpreted as verbs or nouns. It may be possible to produce a variation of COLLPROD that has greater controlled production of collocates and that explicitly exploits the variable types of responses between proficient and less proficient learners. Such an adaptation of COLLPROD could be based on a set of frequent lexical items that can function as both nouns and verbs, as shown in the table below.

Table 11.1 Example items for a future adaptation of COLLPROD

Possible left collocate	Nominal-verbal homonyms	Possible right collocate
WRITE	PLAY	ROLE
HAVE	WASH	CAR
PHYSICAL	EXERCISE	POWER
GIVE	ADDRESS	BOOK

Frequent left and right collocates of these homonyms could be used in multi-word displays where the items were arranged for learners to create collocation links between any two items. Less proficient learners might notice the most obvious collocations links, but not see the hidden layers of links that the some of the nominal-verbal homonyms have. Thus, **address book** might be chosen by most learners, but **give (an) address** might be only noticed by more proficient learners.

It seems to me that developing such an adapted measure of productive collocation knowledge—which goes by the provisional name of COLLINKS—could broadly follow some of design features of recent work into word associations within controlled computerized environments (e.g., Meara & Wolter, 2004; Wolter, 2005). For example,

the display of one set of collocation items might look like the following (see Figure 11.1 below).

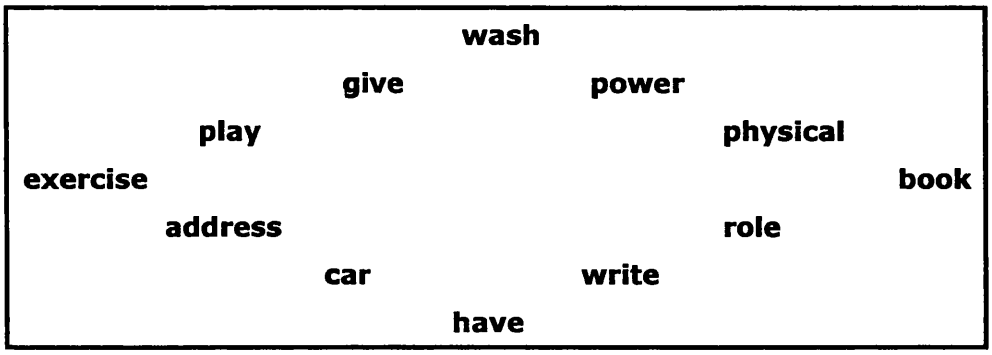

Figure 11.1 Example display of 12 items for linking

Once the items are arranged in a display, other collocation links become possible (depending on the particular set of items). The example set in Figure 11.1 could involve at least 13 possible collocation links, as shown in Figure 11.2 below:

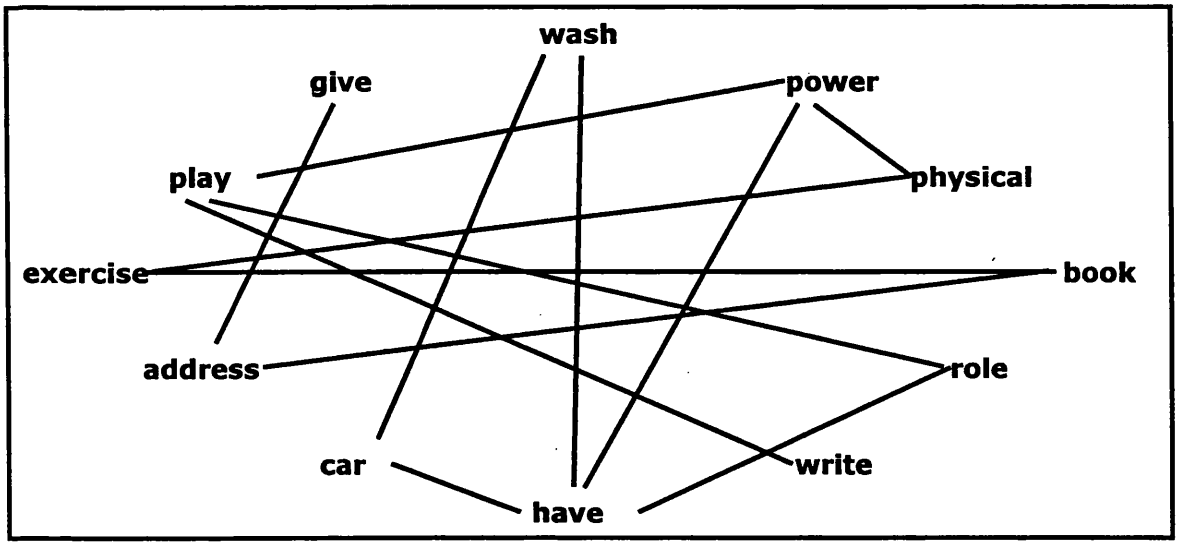

Figure 11.2 Example display of possible collocation links

If each display in COLLINKS involved 10 or more possible collocation links, a test with 9 displays (=90 potential collocations) could produce as much data as COLLPROD does. More interesting, though, is the fact that the question of individual variation would be partially controlled, so the adapted version test might lead to more consistency in the responses that learners gave.

All this is a long way from a contextualized measure of collocation production and from a textual interpretation of L2 collocation knowledge. To design any experimental measure of collocation knowledge necessarily means moving away from a textual interpretation. It also draws us towards a typological view of collocation, which was evident in the collocation recognition experiments that I carried out. Surprisingly, the V+N recognition experiment started from a large database of 99 real collocations, but finished up focusing on just the 15 most recognized collocations. It might be possible to adapt the measure from V+N combinations alone by including Adjectival + Noun and, possibly, Verb + Adverbial items. The knowledge state ranking instrument worked robustly well, but one problem is that we may learn from such an expanded measure little other than lexically congruent collocations are the most strongly known. On the one hand, further confirmation of this (or discovery of other insights) is needed in order to make more confident, empirically derived claims about L2 collocation recognition knowledge. On the other, it is unlikely that such research can help us understand how the L2 lexicon works. In the end, I am not completely convinced such recognition experiments are a useful path to follow, particularly as we found little correlation between recognition knowledge and overall English proficiency.

At the same time, I do believe there is much for us to understand about how learners go about developing their L2 collocation knowledge and how they perceive the

changes that they go through. All three students in the qualitative case study proved to be fascinating: Kensuke's control and resistance, Shoichi's creativity, fluctuation and struggle, and Emi's self-direction and reflective sophistication. The area for exploring the connections between associative knowledge, paraphrase and L2 collocation development would seem to be very promising indeed. Whether such work should be qualitative or not is open to question, but I feel that using measures such as COLLPROD and COLLINKS over time, and conducting longitudinal interviews with a larger group of students, could be a very interesting route to follow. There are, in other words, many possibilities for exploring L2 collocation knowledge further.

To round things off, I hope the experimental work presented in this thesis and the suggestions in this final chapter are of use to a wider community of researchers interested in exploring L2 collocation knowledge. The exploration has been fascinating thus far, but there is still much to do.

References

Aisenstadt, E. (1979). Collocability restrictions in dictionaries. In R.R.K. Hartmann (ed.), *Dictionaries and their users* (pp.71-74).

Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *ITL (Instituut voor Toegepaste Linguistiek) Review of Applied Linguistics*, 53, 53-61.

Aijmer, K. & Altenberg, B. (1991, eds.). *English corpus linguistics: Studies in honour of Jan Svartvik*. London: Longman.

Aitchison, J. (1994, 2nd ed.). *Words in the mind*. Oxford: Blackwell.

Albrechtsen, D., Haastrup, K. & Henriksen, B. (2004, eds.). *Angles on the English-speaking world: writing and foreign language acquisition, 4*. Copenhagen: Museum Tusculanum Press, University of Copenhagen.

Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Allan, Q. (1995). When is a word not a word? Delexical verbs and degrees of desemanticization. *Hong Kong Polytechnic University Working Papers in ELT & Applied Linguistics*, 1(1), 1-14.

Allan, Q. (1998). Delexical verbs and degrees of desemanticization. *Word*, 49(1), 1-17.

Altenberg, B. & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173-195.

Amosova, N. N. (1963). *Osnovui anglijskoy frazeologii*. [Basic English phraseology]. Leningrad: University Press.

Arnaud, P.J.L. & Béjoint, H. (1992, eds.). *Vocabulary and applied linguistics*. London: Macmillan.

Bacon, F. (1627). *Natural history*. In J. Spedding, R. L. Ellis & D. Denon Heath (eds.), *The works of Francis Bacon*. London: Longmans (reprinted, 1875).

Bahns, J. (1993). Lexical collocations: a contrastive view? *ELT Journal*, 47(1), 56-63.

Bahns, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101-114.

Barnes, D. (1976). *From communication to curriculum*. Harmondsworth: Penguin.

Bazell, C.E., Catford, C., Halliday, M.A.K. & Robbins, R.H. (1966, eds.). *In memory of J.R. Firth*. London: Longmans.

Beglar, D. & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing*, 16(2), 131-162.

Béjoint, H. (1994). *Modern lexicography*. Oxford: Oxford University Press.

Benson, M. (1985). Collocations and idioms. In R. Ilson (ed.), *Dictionaries, lexicography and language learning* (pp.61-68).

Benson, M. (1989). The structure of the collocational dictionary. *International Journal of Lexicography*, 2(1), 1-14.

Benson, M. (1990). Collocations and general-purpose dictionaries. *International Journal of Lexicography*, 3(1), 23-34.

Benson, M., Benson, E. & Ilson, R. (1986a). *Lexicographic description of English.* Amsterdam: John Benjamins.

Benson, M., Benson, E. & Ilson, R. (1986b). *The BBI combinatory dictionary of English.* Amsterdam: John Benjamins.

Benson, M., Benson, E. & Ilson, R. (1997). *The BBI dictionary of English word combinations*. Amsterdam: John Benjamins.

Benson, P. (2001). *Teaching and researching autonomy in language learning*. London: Longman.

Benson, P. & Voller, P. (1997, eds.). *Autonomy and independence in language learning.* London: Longman.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken English*. Beijing: Pearson Education/Foreign Language Teaching and Research Press.

Bird, N., Falvey, P., Tsui, A., Allison, D. & McNeil, A. (1994, eds.), *Language and learning*. Hong Kong: Government Printer.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations. A Polish/German empirical study. In P.J.L. Arnaud & H. Béjoint (eds.), *Vocabulary and applied linguistics* (pp. 85-93).

Blue, G. M. & Mitchell, R. (1996, eds.). *Language and education: British studies in applied linguistics 11*. Clevedon: Multilingual Matters.

Bogaards, P. & Laufer, B. (2004, eds.). *Vocabulary in a second language*. Amsterdam: John Benjamins.

Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model: fundamental measurement in the human sciences*. New Jersey: Lawrence Erlbaum.

Bonk, W. J. (2000). Testing ESL learners' knowledge of collocations. *Educational Resources Information Center Research Report ED 442 309* (69 pp.).

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

Brinton, L.J. & Akimoto, M. (1999, eds.). *Collocational and idiomatic aspects of composite predicates in the history of English*. Amsterdam: John Benjamins.

Brown, G., Malmkjaer, K. & Williams, J. (1996, eds.). *Performance and competence in*

second language acquisition. Cambridge: Cambridge University Press.

Camic, P., Rhodes, J. & Yardley, L. (2000, eds.). *Qualitative research in psychology.* Washington, DC: American Psychological Association.

Carter, R. (1998, 2nd ed.). *Vocabulary: applied linguistic perspectives.* London: Routledge.

Carter, R. (2004). Introduction. In J. Sinclair, *Trust the text* (pp.1-6).

Carter, R. & McCarthy, M.J. (1988, eds.). *Vocabulary and language teaching.* London: Longman.

Chi Man-lai, Wong Pui-yiu, & Wong Chau-ping (1994). Collocational problems amongst ESL learners: a corpus-based study. In L. Flowerdew & A.K.K Tong (eds.), *Entering text* (pp. 157-163).

Coady, J. & Huckin, T. (1997, eds.). *Second language vocabulary acquisition.* Cambridge: Cambridge University Press.

Cobb, T. (2003). *Web VP* (v1.5). Available at: http://www.er.uqam.ca/nobel/r21270/cgi-bin/webfreqs/web_vp.cgi

Cole, M., John-Steiner, V., Scribner, S. & Souberman, E. (1978, eds.). *L.S. Vygotsky Mind in society: the development of higher psychological processes.* Cambridge, Massachusetts: Harvard University Press.

Collins COBUILD Advanced Learner's English Dictionary (2006, 5th edition). Glasgow: HarperCollins.

Collins COBUILD English Collocations on CD-rom v1.1 (1995a). Glasgow: HarperCollins.

Collins COBUILD English Language Dictionary (1987, 1st ed.). Glasgow: HarperCollins.

Collins COBUILD English Dictionary (1995b, 2nd ed.). Glasgow: HarperCollins.

Collins COBUILD New Student's Dictionary (1997). Glasgow: HarperCollins.

Collins COBUILD Student's Dictionary (1987). Glasgow: HarperCollins.

Collins WordbanksOnline (2004). London: HarperCollins. Available at: http://www.collins.co.uk/Corpus/CorpusSearch.aspx

Coulthard, M. (1994, ed.). *Advances in written text analysis.* London: Routledge.

Cowie, A.P. (1978). The place of illustrative material and collocations in the design of a learner's dictionary. In P. Strevens (ed.), *In honour of A.S. Hornby* (pp.127-139).

Cowie, A.P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223-235.

Cowie, A.P. (1998a). Introduction. In A.P. Cowie (ed.), *Phraseology: theory, analysis, and applications* (pp.1-20).

Cowie, A.P. (1998b). Phraseological dictionaries: some East-West comparisons. In A.P.

Cowie (ed.), *Phraseology: theory, analysis, and applications* (pp.209-228).

Cowie, A.P. (1998c, ed.). *Phraseology: theory, analysis, and applications*. Oxford: Oxford University Press.

Cowie, A.P. (1999). *English dictionaries for foreign learners*. Oxford: Oxford University Press.

Cowie, A.P. & Mackin, R. (1975). *Oxford dictionary of current idiomatic English Volume 1. Verbs with prepositions and particles*. Oxford: Oxford University Press.

Cowie, A.P. & Howarth, P. (1996). Phraseological competence and written proficiency. In G.M. Blue & R. Mitchell (eds.), *Language and education: British studies in applied linguistics 11* (pp. 80-93).

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly,* 34(2), 213-238.

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics*, 11(2), 183-199.

Dale, E. (1965). Vocabulary measurement: Technique and major findings. *Elementary English,* 42, 895-901, 948.

Dale, P.S. (1972, 2nd ed.). *Language development structure and function*. New York: Holt, Rinehart and Winston.

Dam, L. (1991). *Learner autonomy: from theory to practice*. Dublin: Authentik.

de Groot, A.M.B. & Kroll, J.F. (1997, eds.). *Tutorials in bilingualism*. Mahwah, New Jersey: Lawrence Erlbaum.

de Saussure, F. (1915). *Course in general linguistics.* New York: McGraw-Hill. [Edited by C. Bally & A. Sèchehaye. Translated by W. Baskin.]

Dechert, H. W. & Lennon, P. (1989). Collocational blends of advanced learners: a preliminary analysis. In W. Olesky (ed.), *Contrastive pragmatics* (pp.131-168).

Denzin, N. & Lincoln, Y. (2003, eds.). *Collecting and interpreting qualitative materials* (2nd ed.). London: Sage.

Dromi, E. (1987). *Early lexical development*. San Diego: Singular Publishing Group.

Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliffs, New Jersey: Prentice-Hall.

Ellis, N.C. (1994a, ed.). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, N.C. (1994b). Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation. In N.C. Ellis (ed.), *Implicit and explicit learning of languages* (pp.211-282).

Ellis, N.C. & Beaton, A. (1995). Psycholinguistic determinants of foreign language vocabulary learning. In B. Harley (ed.), *Lexical issues in language learning* (pp.107-165).

Farghal, M. & Obiedat, H. (1995). Collocations: a neglected variable in EFL. *International Journal of Applied Linguistics*, 28(4), 313-331.

Firth, J. R. (1952/3). Linguistic analysis as a study of meaning. In F. R. Palmer (ed.), *Selected papers of J.R. Firth* (pp.12-26).

Firth, J. R. (1956). Descriptive linguistics and the study of English. In F. R. Palmer (ed.), *Selected papers of J.R. Firth* (pp.96-113).

Firth, J. R. (1957a). A synopsis of linguistic theory. 1930-55. In F. R. Palmer (ed.), *Selected papers of J.R. Firth* (pp.168-205).

Firth, J. R. (1957b). Ethnographic analysis and language with reference to Malinowoski's views. In F. R. Palmer (ed.), *Selected papers of J.R. Firth* (pp.137-167).

Firth, J. R. (1957c). Modes of meaning. In F. R. Palmer (ed.), *Papers in linguistics 1934-1951* (pp.190-215).

Flowerdew, L. & Tong, A.K.K. (1994, eds.). *Entering text*. Hong Kong: Hong Kong University of Science and Technology Language Centre.

Fox, G. (1987). The case for examples. In J. Sinclair (ed.), *Looking up: an account of the COBUILD project in lexical computing* (pp.137-149).

Gabryś-Biskup, D. (1990). *Some remarks on combinability: lexical collocations*. Katowice: University of Silesia.

Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S. Kuczaj (ed.), *Language development, volume 2: language, thought and culture* (pp.301-334).

Gitsaki, C. (1996). The development of ESL collocational knowledge. *Unpublished PhD thesis*. University of Queensland.

Granger, S. (1998a). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (ed.), *Learner English on the computer* (pp.3-18).

Granger, S. (1998b, ed.). *Learner English on the computer*. London: Longman.

Granger, S. (1998c). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Cowie, A.P. (ed.), *Phraseology: theory, analysis, and applications* (pp. 145-160).

Gu Yongqi (1994). Vocabulary learning strategies of good and poor Chinese EFL learners. In N. Bird, P. Falvey, A. Tsui, D. Allison & A. McNeil (eds.), *Language and learning* (pp. 376-401).

Gu Yongqi, P. (2003). Fine brush and freehand: the vocabulary-learning art of two successful Chinese EFL learners. *TESOL Quarterly*, 37(1), 73-104.

Gu Yongqi & Johnson, K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46(4), 643-679.

Halliday, M.A.K. (1966). Lexis as a linguistic level. In C.E. Bazell, C. Catford, M.A.K. Halliday & R.H. Robbins (eds.), *In memory of J.R. Firth* (pp.148-162).

Halliday, M.A.K. (1985). *An introduction to functional grammar*. London: Arnold.

Halliday, M.A.K. & R. Hasan (1976). *Cohesion in English*. London: Longman.

Hammerberg, B. (1993, ed.). *Problems, process and product in language learning*. Åbo, Finland: Åbo Academic.

Hartmann, R.R.K (1979, ed.). *Dictionaries and their users*. Exeter: Exeter University.

Hausmann, F.J. (1979). Un dictionnaire des collocations est-il possible? *Travaux de linguistique et de literature*, 17, 187-195.

Heim, A.W. & Watts, K.P. (1961). A preliminary study of the self-judging vocabulary scale. *British Journal of Psychology*, 52(2), 175-186.

Henwood, K. & N. Pidgeon (2003). Grounded theory in psychological research. In P. Camic, J. Rhodes & L. Yardley (eds.), *Qualitative research in psychology* (pp. 131-151).

Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.

Hoey, M. (2001). *Textual interaction*. London: Routledge.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.

Hornby, A.S. (1946). Linguistic pedagogy I. The doctrines of de Saussure. contextual procedure—word families. *English Language Teaching*, 1(1), 7-11.

Hornby, A.S. (1947a). Editorial: vocabulary control and vocabulary layout. *English Language Teaching*, 1(4), 89-90.

Hornby, A.S. (1947b). Linguistic pedagogy IV. The contextual procedure—word families. *English Language Teaching*, 1(4), 91-95.

Hornby, A.S., Gatenby, E.V. & Wakefield, H. (1942). *Idiomatic and syntactic English*. Tokyo: Kaitakusha.

Hornby, A.S., Gatenby, E.V. & Wakefield, H. (1948). *A learner's dictionary of current English*. Oxford: Oxford University Press [Retitled in 1952 as *The advanced learner's dictionary of current English*].

Howarth, P. (1998). The phraseology of learners' academic writing and second language proficiency. In A.P. Cowie (ed.), *Phraseology: theory, analysis, and applications* (pp. 161-186).

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hulstijn, J. (1997). Mnemonic methods in foreign language vocabulary learning: theoretical considerations and pedagogical implications. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition* (pp. 203-224).

Hunston, S. & Francis, G. (2000). *Pattern grammar*. Amsterdam: John Benjamins.

Hunt, K.W. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732-739.

Ilson, R. (1985, ed.). *Dictionaries, lexicography and language learning*. Oxford: The British Council/Pergamon Press.

Jeffrey, G. B. (1952). Foreword. In M. West, *A general service list of English words* (pp.v-vi).

Jespersen, O. (1933). *Essentials of English grammar*. New York: Henry Holt.

Jiang, N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language* (pp.101-126).

Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, 11(2), 149-158.

Johnson, S. (1755/1967). *A dictionary of the English language*. London: Strahan / New York: AMS Press.

Jones, S. & Sinclair, J. (1974). English lexical collocations: A study in computational linguistics. *Cahiers de Lexicologie*, 24, 15-61.

Kamimoto, T. (2001, 24 November). An examination of Nation's (1990) vocabulary levels test. *Paper presented at JALT/PAC 3*. Kitakyushu, Japan.

Kilgarriff, A. (March, 1996). *BNC database and word frequency lists*. Available at: http://www.itri.bton.ac.uk/~Adam.Kilgarriff/bnc-readme.html

Kjellmer, G. (1991). A mint of phrases. In K. Aijmer & B. Altenberg (eds.), *English corpus linguistics* (pp.111-127).

Kojić-Sabo, I. & Lightbown, P. (1999). Students' approaches to vocabulary learning and their relationship to success. *The Modern Language Journal*, 83(2), 176-192.

Krishnamurty, R. (2004, ed.). *English collocation studies: the OSTI report.* London: Continuum.

Kroll, J. (1993). Accessing conceptual representations for words in a second language. In R. Shreuder & B. Weltens (eds.), *The bilingual lexicon* (pp. 53-81).

Kroll, J.F. & de Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual: mapping form to meaning in two languages. In A.M.B. de Groot & J.F. Kroll (eds.), *Tutorials in bilingualism* (pp.169-199).

Kuczaj, S. (1982, ed.). *Language development, volume 2: language, thought and culture*. Hillsdale, New Jersey: Lawrence Erlbaum.

Kunnan, A.J. (1998, ed.), *Validation in language assessment*. Mahwah, New Jersey: Lawrence Erlbaum.

Lennon, P. (1996). Getting 'easy' verbs wrong at the advanced level. *IRAL*, 34, 23-36.

Lewis, M. (2000). *Teaching collocation*. Hove: Language Teaching Publications.

Linacre, J. M. (2005). *WINSTEPS* (Version 3.57). Chicago, Illinois.

Little, D. (1991). *Learner autonomy: definitions, issues and problems*. Dublin: Authentik.

Little, D. (1997). Responding authentically to authentic texts. In P. Benson & P. Voller (eds.), *Autonomy and independence in language learning* (pp.225-248).

Longman Dictionary of Contemporary English (1995, 3rd ed.). Harlow: Longman.

Marton, W. (1977). Foreign vocabulary learning as problem No.1 of language teaching at the advanced level. *Interlanguage Studies Bulletin*, 2, 33-57.

McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.

McNamara, T. (1996). *Measuring second language performance*. London: Longman.

Meara, P.M. (1989). Matrix models of vocabulary acquisition. In P. Nation, P. & R. Carter (eds.), *Vocabulary acquisition*, AILA Review-Revue de l'AILA, 6.

Meara, P.M. (1992). Network structures and vocabulary acquisition in a foreign language. In P.J.L. Arnaud & H. Béjoint (eds.), *Vocabulary and applied linguistics* (pp.62-70).

Meara, P.M. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (eds.), *Performance and competence in second language acquisition* (pp.35-53).

Meara, P.M. (1999, November; revised). The vocabulary knowledge framework. *Vocabulary Acquisition Research Group Discussion Paper*. Available at: http://www.swan.ac.uk/cals/calsres/vlibrary/pm96d.htm

Meara, P.M. & Buxton, T. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142-154.

Meara, P.M. & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19-30.

Meara, P.M. & Milton, J. (2003). *X_Lex: the Swansea vocabulary levels test*. Newbury: Express.

Meara, P.M. & Wolter, B. (2004). V_Links: Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup & B. Henriksen (eds.), *Angles on the English-speaking world: writing and foreign language acquisition* (pp.85-96).

Mel'čuk, I. (1998). Collocations and lexical functions. In A.P. Cowie (ed.) *Phraseology: theory, analysis, and applications* (pp.24-53).

Melka Teichroew, F.J. (1982). Receptive versus productive vocabulary: a survey. *Interlanguage Studies Bulletin*, 6(2), 5-33.

Miller, G.A. & Fellbaum, C. (1991). Semantic networks of English. *Cognition*, 41, 197-229.

Mitchell, T. F. (1971). Linguistic 'goings on': collocations and other lexical matters arising on the syntagmatic record. *Archivum Linguisticum*, 2, 35-69.

Mochizuki, M. (1998). A vocabulary size test for Japanese learners of English. *The IRLT Bulletin*, 12, 27-53.

Mochizuki, M. (2002). Exploration of two aspects of vocabulary knowledge: Paradigmatic and collocational. *Annual Review of English Language Education in Japan*, 13, 121-129.

Moon, R. (1987). The analysis of meaning. In J. Sinclair (ed.), *Looking up: an account of the COBUILD project in lexical computing* (pp.86-103).

Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford: Oxford University Press.

Nation, P. & Carter, R. (1989, eds.). *Vocabulary acquisition*. AILA Review-Revue de l'AILA, 6.

Nation, I.S.P. (1990). *Teaching and learning vocabulary*. Boston, Mass: Heinle & Heinle.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I.S.P. & Kyongho, H. (1995). Where would general service vocabulary stop and special purpose vocabulary begin? *System*, 23, 35-41.

Nattinger, J.R. & DeCarrico, J.S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nelson, D.L. & Schreiber, T.A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language*, 31, 237-260.

Neselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.

Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Olesky, W. (1989, ed.). *Contrastive pragmatics*. Amsterdam/Philadelphia: John Benjamins.

Oller, J.W. (1979). *Language tests at school*. London: Longman.

Oxford Advanced Learner's Dictionary (1974, 3rd ed.). Oxford: Oxford University Press.

Oxford Advanced Learner's Dictionary (1989, 4th ed.). Oxford: Oxford University Press.

Oxford Advanced Learner's Dictionary (2000, 6th ed.). Oxford: Oxford University Press.

Oxford Collocations Dictionary for Students of English (2002). Oxford: Oxford University Press.

Palmer, F.R. (1957a, ed.). *Selected papers in linguistics 1934-1951*. London: Oxford University Press.

Palmer, F.R. (1957b, ed.). *Selected papers of J.R. Firth*. London: Longmans.

Palmer, H. (1929a). What shall we call "a word"? *The Bulletin of the Institute for Research in English Teaching*, 54, 1-2.

Palmer, H. (1929b). What is an idiom? *The Bulletin of the Institute for Research in English Teaching*, 56, 1-2.

Palmer, H. (1931). *First interim report on vocabulary selection*. Tokyo: Kaitakusha.

Palmer, H. (1933a). Our research on collocations. *The Bulletin of the Institute for Research in English Teaching*, 95, 1-2.

Palmer, H. (1933b). *Second interim report on English collocations*. Tokyo: Kaitakusha.

Palmer, H. (1934). Director's report. *The Bulletin of the Institute for Research in English Teaching*, 108, 16-24.

Paribakht, T. S. & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition* (pp.174-200).

Pawley, A. & Syder, F.H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schmidt (eds.), *Language and communication* (pp.191-225).

Philip, G. (2004). Habeas corpus: direct access, salience and delexicalisation in corpus-based metaphor studies. *Paper presented at Mind, Language and Metaphor: EURESCO Conference on the Processing of Metaphor and Metonymy—From Computers to Neuropsychology*. Granada, Spain.

Read, J. (1993). The development of a new measure depth of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A.J. Kunnan (ed.), *Validation in language assessment* (pp. 41-60).

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: how should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language* (pp.209-227).

Renouf, A. (1987). Moving on. In J. Sinclair (ed.), *Looking up: an account of the COBUILD project in lexical computing* (pp.167-178).

Renouf, A. & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (eds.), *English corpus linguistics: studies in honour of Jan Svartvik* (pp. 128-143).

Richards, J.C. & Schmidt, R.W. (1983, eds.), *Language and communication*. London: Longman.

Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.

Saito, H. (1915). [Jukugo honi ei-wa chu-jiten] *Saito's idiomological English-Japanese dictionary*. Tokyo: Iwanami.

Scarcella, R. & Zimmerman, C. (1998). Academic words and gender: ESL student performance on a test of academic lexicon. *Studies in Second Language Acquisition*, 20, 27-49.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (eds.), *Vocabulary description, acquisition and pedagogy* (pp. 199-227).

Schmitt, N. (1998a). Measuring collocation knowledge: key issues and an experimental assessment procedure. *ITL Review of Applied Linguistics,* 27-47, 119-120.

Schmitt, N. (1998b). Tracking the incremental acquisition of second language vocabulary: a longitudinal study. *Language Learning*, 48, 281-317.

Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing*, 16(2), 189-216.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. & McCarthy, M. (1997, eds.). *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Schmitt, N. & Meara, P.M. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition*, 19, 17-36.

Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.

Scott, M.R. (1998). *Wordsmith tools*. Oxford: Oxford University Press.

Shillaw, J. (2001, July). Scaling yes / no tests. *Paper presented at the 11th Vocabulary Acquisition Research Group Network Conference, Swansea University, UK.*

Shiotsu, T. (2001). An analysis of the Vocabulary Levels Test. *Bulletin of the Institute of Foreign Language Education Kurume University*, 8, 27-49.

Shreuder, R. & Weltens, B. (1993, eds.). *The bilingual lexicon*. Amsterdam: John Benjamins.

Silverman, D. (2000). *Doing qualitative research*. London: Sage.

Silverman, D. (2003). Analyzing talk and text. In N. Denzin & Y. Lincoln (eds.), *Collecting and interpreting qualitative materials* (2nd ed., pp. 340-362). London: Sage.

Sinclair, J. (1966). Beginning the study of lexis. In C.E. Bazell, C. Catford, M.A.K. Halliday & R.H. Robbins (eds.), *In memory of J.R. Firth* (pp.410-430).

Sinclair, J. (1985). Lexicographic evidence. In R. Ilson (ed.), *Dictionaries, lexicography and language learning* (pp.81-94).

Sinclair, J. (1987a). Collocation: a progress report. In R. Steele & T. Threadgold (eds.), *Language topics: Essays in honour of Michael Halliday* (pp.319-331).

Sinclair, J. (1987b, ed.). *Looking up: an account of the COBUILD project in lexical computing*. London: HarperCollins.

Sinclair, J. (1987c). The nature of the evidence. In J. Sinclair (ed.), *Looking up: an account of the COBUILD project in lexical computing* (pp.150-159).

Sinclair, J. (1990a, ed.). *Collins COBUILD English grammar*. London: Collins.

Sinclair, J. (ed.) (1990b). *COBUILD English learner's dictionary*. London: Collins / University of Birmingham.

Sinclair, J. (1991). *Corpus concordance collocation*. Oxford: Oxford University Press.

Sinclair, J. (1994). Trust the text. In M. Coulthard (ed.), *Advances in written text analysis* (pp.12-25).

Sinclair, J. (2004a). *Trust the text*. London: Routledge.

Sinclair, J. (2004b). The lexical item. In *Trust the text* (pp.131-148).

Sinclair, J. (2004c). The search for units of meaning. In *Trust the text* (pp. 24-48).

Sinclair, J., Jones, S. & Daley, R. (1970). *English lexical studies: Report to OSTI on Project C/LP/08*. University of Birmingham: Department of English.

Sinclair, J.M. & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M.J. McCarthy (eds.), *Vocabulary and language teaching* (pp.140-158).

Skehan, P. (1989). *Individual differences in second-language learning*. London: Arnold.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Smith, R.C. (1998). The Palmer-Hornby contribution to English teaching in Japan. *International Journal of Lexicography*, 11(4), 269-291.

Söderman, T. (1993). Word associations of foreign language learners and native speakers—different response types and their relevance to lexical development. In B. Hammerberg (ed.), *Problems, process and product in language learning* (pp.157-169).

Steele, R. & Threadgold, T. (1987, eds.). *Language topics: essays in honour of Michael Halliday*. Amsterdam: John Benjamins.

Stein, G. (1991). The phrasal verb type 'to have a look' in modern English. *International Review of Applied Linguistics 19*(1), 1-21.

Stein, G. & Quirk, R. (1991). On having a look in a corpus. In Aijmer, K. & B. Altenberg (eds.), *English corpus linguistics* (pp.197-203).

Strässler, J. (1982). *Idioms in English: a pragmatic analysis*. Tűbingen: Gunter Narr Verlag.

Strauss, A. & Corbin, J. (1998, 2nd ed.). *Basics of qualitative research*. London: Sage.

Strevens, P. (1978, ed.). *In honour of A.S. Hornby*. Oxford: Oxford University Press.

Stubbs, M. (2001). *Words and phrases*. Oxford: Blackwell.

Swales, J.M. & Feak, C.B. (1994). *Academic writing for graduate students*. Ann Arbor: University of Michigan Press.

Sweet, H. (1891). *A new English grammar (part 1)*. Oxford: Clarendon Press.

Tanabe, H. (1999). Composite predicates and phrasal verbs in The Paston Letters. In L.J. Brinton & M. Akimoto (eds.), *Collocational and idiomatic aspects of composite predicates in the history of English* (pp.97-132).

Teubert, W. (2004). Interview with John Sinclair. In R. Krishnamurty (ed.), *English collocation studies: The OSTI report* (pp.xvii-xxix).

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Trager, G. (1940). The Russian gender categories. *Language*, 16, 300-307.

Vinogradov, V. V. (1947). Ob osnovnuikh tipakh frazeologicheskikh editnits v russkom yazuike. [On basic typical phraseological units in the Russian language]. In *A.A. Shakhmatov, 1864-1920. Sbornik statey i materialov* (pp.339-364). Moscow: Nauka.

Vygotsky, L. (1935). *Interaction between learning and development.* In M. Cole, V. John-Steiner, S. Scribner & E. Souberman (eds.), *L.S. Vygotsky. Mind in society: the development of higher psychological processes* (pp.79-91).

Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. *Occasional Papers of Notre Dame Seishin University*. Available at: http://www1.harnet.ne.jp/~waring/papers/scales.htm.

Wenden, A. (1991). *Learner strategies for learner autonomy*. New York: Prentice Hall.

Wesche, M. & Paribakht, T.S. (1996). Assessing L2 vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.

West, M. (1953, ed.). *A general service list of English words*. London: Longman.

Wierzbicka, A. (1982). Why can you have a drink when you can't *have an eat? *Language*, 58(4), 753-799.

Wolter, B. (2002). Assessing proficiency through word associations: is there still hope? *System*, 30(3), 315-329.

Wolter, B. (2005). V_Links: a new approach to assessing depth of word knowledge. *Unpublished PhD thesis*. University of Swansea, Wales.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Zhang, Xiaolin (1993). English collocations and their effect on the writing of native and non-native college freshmen. *Unpublished PhD thesis*. Indiana University of Pennsylvania.