

**A Corpus-Based Error Analysis of Korean Learner English: From a
Cognitive Linguistic Perspective to the L2 Mental Lexicon**

by

WONKYUNG CHOI

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Department of English Language
and Linguistics
College of Arts and Law
The University of Birmingham
January 2019

**UNIVERSITY OF
BIRMINGHAM**

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis investigates lexical errors in a learner corpus that consists of essays written by Korean learners of English. Taking a cognitive linguistic perspective, it first presents the L2 lexical development model as a conceptual framework. Then, based on this framework, it proposes a new error taxonomy in which errors from the four lexical domains in the L2 mental lexicon can be categorised as either interlingual or intralingual errors. In order to identify whether the taxonomy is well-grounded, this thesis selects four error features, one from each of the four lexical domains: collocational errors of dimensional adjectives in the semantic domain; over-passivisation errors of non-alternating unaccusative verbs in the syntactic domain; derivational morphological errors in the morphological domain; spelling errors in the phonological/orthographic domain. The results, obtained through corpus-based error analysis, suggest that there is evidence of both interlingual and intralingual influences on the four error features from the four lexical domains. Based on the findings, this study provides pedagogical implications for English classrooms in Korea and recommends that the findings should be used to improve teaching materials and to raise awareness of the influences on the lexical errors.

Acknowledgements

I am deeply grateful to all the people who have helped make the completion of this thesis possible. Without them, it would not have been possible for me to walk this journey.

First and foremost I would like to thank from the bottom of my heart my lead supervisor Dr Crayton Walker, for his commitment, guidance and constant encouragement during my research at Birmingham.

I would also like to express my profound gratitude to my supervisors in South Korea, Professor Inn-Chull Choi, Professor Dong Ju Lee, and Professor Kyung-Sung Kim. It would not have been possible even to start my PhD without their support.

My special thanks also go to my two examiners, Professor Fanny Meunier and Dr Paul Thompson, who read my thesis thoroughly and gave me critical, but constructive feedback. I am thankful to Ann Marie Wareham who has carefully proofread this thesis.

I am also deeply indebted to my friends, who have supported me with loving kindness. They are: Pastor Min Lee, Pastor Jangyob Kim, Deacon Sang Sun Ju, Deacon Suryun Park and Changwook Kim.

However, my deepest gratitude will always be to my wife Minjeong Jo (Grace) and my lovely sons Kyungmin Choi (Joshua) and Jimin Choi (James), who have encouraged me to try to become a better husband and father. I would also like to thank my parents (Hyontac Choi & Yangsoo Chung) and my mother-in-law (Myungja Yun) for their unconditional love and support.

Finally, I dedicate this thesis to God, who always loves me and guides me on life's path.

Contents

CHAPTER 1: INTRODUCTION	1
1.1 Overview of the Thesis	1
1.2 Rationale of the Study	4
1.2.1 The Significance of Learners' Errors for English Language Teaching (ELT)	4
1.2.2 Pedagogic Value of the Study	6
1.3 Objectives of the Study	8
1.4 Organisation of the Thesis	9
CHAPTER 2: ERROR ANALYSIS, COGNITIVE LINGUISTIC PERSPECTIVE TO L2 LEARNERS' ERRORS AND LEARNER CORPUS RESEARCH	11
2.1 Theoretical Background of Error Analysis	11
2.1.1 Error Analysis and Language Teaching and Learning	11
2.1.2 Error Categorisation Procedures	14
2.1.3 Historical Background of EA Studies: Two Phases	16
2.1.4 The First Phase: Review of EA Studies during the EA Movement.....	17
2.1.5 The Second Phase: Lexical Error and its Taxonomies	21
2.1.6 A Critical Issue of EA: Errors vs. Mistakes.....	26
2.1.7 This Study's Approach to Lexical Errors.....	28
2.2 Cognitive Linguistic Perspective to Learners' Errors	32
2.2.1 Cognitive Linguistic Perspective to Language Acquisition.....	32
2.2.2 Contributions of the Cognitive Linguistic Perspective to this Study	36
2.3 Theoretical Background of Learner Corpus Research	37
2.3.1 Corpus Linguistics and Learner Corpus Research.....	37
2.3.2 Learner Corpus Research Methodology: CIA and CEA	40
2.3.3 Corpus-based Research on Korean Learners of English in Korea	43
2.4 Chapter Summary	45
CHAPTER 3: LANGUAGE PRODUCTION MODEL AND A NEW ERROR TAXONOMY	47
3.1 Overview of the Mental Lexicon in the Human Mind	48
3.2 Theoretical Issues in the Models of the L2 Mental Lexicon	51

3.2.1 Single (Shared) vs. Separate Mental Lexicon	52
3.2.2 Revised Hierarchical Model (RHM)	54
3.2.3 Language Selective vs. Language Non-selective Hypotheses	57
3.2.4 Perspectives of the Current Study toward the L2 Mental Lexicon.....	60
3.3 Language Production Model	64
3.3.1 Levelt's L1 Speech Production Model	65
3.3.2 De Bot's Multilingual Production Model.....	67
3.3.3 Jiang's L2 Vocabulary Acquisition Model	70
3.3.4 Contributions of the Three Language Production Models to the Current Study	74
3.4 Sources of Lexical Errors	75
3.4.1 Diagnostic Approach to Lexical Errors	75
3.4.2 Sources of Errors and the Role of Cognitive Vehicles.....	79
3.5 The L2 Lexical Development Model and the New Error Taxonomy	81
3.5.1 L2 Lexical Development Model in the Current Study	81
3.5.2 A New Categorisation Based on Multi-faceted Criteria	84
3.6 Chapter Summary	87
 CHAPTER 4: ERROR FEATURES: SEMANTIC AND SYNTACTIC DOMAINS	89
4.1 Semantic Domain: Collocational Errors of Dimensional Adjectives	90
4.1.1 Collocational Errors Made by L2 Learners.....	92
4.1.2 Dimensional Adjectives and the Possible Sources of the Errors	94
4.1.3 Figurative Extension and Overextension	96
4.1.4 Different Viewpoints Regarding Lexical Errors in the Semantic Domain	97
4.1.5 Corpus-based Approach to Detect Collocational Errors of Dimensional Adjectives	100
4.2 Syntactic Domain: Over-passivisation Errors of Unaccusative Verbs.....	102
4.2.1 The Possible Causes of Over-passivisation Errors of Unaccusative Verbs	104
4.2.2 The Interlingual Influence: L1 Morphology	107
4.2.3 Korean Morphological System of Passives and Causatives	110
4.2.4 The Intralingual Influence: Subject Animacy	112
4.3 Chapter Summary	114
 CHAPTER 5: ERROR FEATURES: MORPHOLOGICAL AND PHONOLOGICAL/ORTHOGRAPHIC DOMAINS	115

5.1 Morphological Domain: Derivational Morphological Errors	116
5.1.1 Metalinguistic Awareness and Morphological Awareness	116
5.1.2 Three Aspects of Derivational Morphology	119
5.1.3 The Effect of Morphological Awareness on L2 Acquisition	121
5.1.4 The Intralingual and Interlingual Influences on the L2 Derivational Morphology	123
5.2 Phonological/Orthographic Domain: Spelling Errors	125
5.2.1 Spelling Errors Made by L2 Learners	126
5.2.2 Sources of Spelling Errors Produced by L2 Learners	127
5.2.3 The Interlingual and Intralingual Spelling Errors Made by Korean Learners of English	131
5.3 Chapter Summary	136
 CHAPTER 6: METHODOLOGY	138
6.1 Research Questions	139
6.2 Corpora	142
6.2.1 The Yonsei English Learner Corpus (YELC)	143
6.2.2 The Bank of English (BoE): the Main English Reference Corpus	147
6.2.3 The British National Corpus (BNC) and <i>MorphoQuantics</i> : Supplementary English Reference Corpora	149
6.2.4 The Sejong Corpus: the Korean Reference Corpus.....	150
6.3 Research Procedures.....	150
6.3.1 Research Procedures: Collocational Errors of Dimensional Adjectives	153
6.3.2 Research Procedures: Over-passivisation Errors of Non-alternating Unaccusative Verbs	156
6.3.3 Research Procedures: Derivational Morphological Errors	158
6.3.4 Research Procedures: Orthographic (Spelling) Errors	162
6.4 Chapter Summary	164
 CHAPTER 7: COLLOCATIONAL ERRORS OF DIMENSIONAL ADJECTIVES	166
7.1 Outline of the Corpus-based Error Analysis for Collocational Errors of Dimensional Adjectives	166
7.2 Interlingual Influence (L1 Transfer): A Possible Source of Collocational Errors of Dimensional Adjectives.....	170
7.3 Positive Interlingual Influence (Positive L1 Transfer).....	178

7.4 Intralingual Influence: Another Possible Source of Collocational Errors of Dimensional Adjectives	181
7.5 Conclusion	186
CHAPTER 8: OVER-PASSIVISATION ERRORS OF NON-ALTERNATING UNACCUSATIVE VERBS.....	188
8.1 Three Types of Over-passivisation Errors of Non-alternating Unaccusative Verbs	188
8.2 Interlingual Influence: L1 Morphology.....	191
8.3 Interlingual Influence According to Proficiency Level.....	196
8.4 Interlingual Influence of Individual Verbs.....	199
8.5 Intralingual Influence: Subject Animacy Effect	201
8.6 Intralingual Influence According to Proficiency Level.....	205
8.7 Intralingual Influence of Individual Verbs.....	207
8.8 Conclusion	209
CHAPTER 9: DERIVATIONAL MORPHOLOGICAL ERRORS.....	211
9.1 Overview of the Corpus-based Error Analysis on Derivational Morphological Errors	211
9.2 The Results According to Proficiency Level	216
9.3 Derivational Morphological Error Types According to the Degree of Overgeneralisation	220
9.4 Intralingual Influence: Overgeneralisation of Affixation Rules	227
9.5 Interlingual Influence Shown in Type D Errors	228
9.6 Conclusion	231
CHAPTER 10: ORTHOGRAPHIC ERRORS	233
10.1 Overview of the Corpus-based Error Analysis on Spelling Errors	234
10.2 Examples of Interlingual and Intralingual Spelling Errors	236
10.3 Interlingual Spelling Errors According to Proficiency Level.....	240
10.4 Intralingual Spelling Errors According to Proficiency Level	244
10.5 Individual Influence of Each Subcategory	247
10.6 Conclusion	251
CHAPTER 11: CONCLUSION	254

11.1 Main Findings of the Study.....	254
11.2 Pedagogic Implications for English Language Teaching	257
11.2.1 Pedagogic Implications for Collocational Errors of Dimensional Adjectives	258
11.2.2 Pedagogic Implications for Non-alternating Unaccusative Verb Errors.....	259
11.2.3 Pedagogic Implications for Derivational Morphological Errors	259
11.2.4 Pedagogic Implications for Orthographic Errors.....	260
11.3 Strengths of the Study	262
11.3.1 The Value of Authentic, Naturally Occurring Data and of Corpus-based Error Analysis	263
11.3.2 Cognitive Linguistic Perspective on L2 Learners' Errors.....	264
11.3.3 Implications of Corpus-based Error Analysis for Other Research	266
11.4 Limitations of the Study.....	267
11.5 Suggestions for Future Research	269
References	272
 Appendices	295
Appendix 1: The general information of the YELC	295
Appendix 2: The data samples of sub-corpora in the YELC	296
Appendix 3: The additional information for collocational errors of dimensional adjectives	299
Appendix 4: The additional information for over-passivisation errors of non-alternating unaccusative verbs	301
Appendix 5: The additional information for derivational morphological errors	303
Appendix 6: The additional information for orthographic errors	306

List of Tables

Table 2.1 List of recent learner corpus research in Korea	44
Table 3.1 The new error taxonomy according to lexical domains and sources	85
Table 4.1 Outline of theoretical background of the error features: semantic and syntactic domains.....	90
Table 5.1 Outline of theoretical background of the error features: morphological and phonological/orthographic domains.....	115
Table 5.2 Comparison between English and Korean consonant systems.....	134
Table 6.1 Numbers of words and texts in the YELC according to proficiency level.....	144
Table 6.2 The sub-corpora of the Bank of English.....	148
Table 6.3 Outline of the research of the four error features.....	152
Table 6.4 The frequency of the associations of <i>large</i> and <i>damage</i> with different slots in the BoE.....	154
Table 6.5 The top ten pre-modifying adjectives of the noun <i>damage</i> by <i>t</i> -score.....	155
Table 6.6 Matched and mismatched non-alternating verbs.....	157
Table 6.7 The levels of affixes based on Bauer & Nation (1993).....	159
Table 6.8 Derivational affixes extracted from the YELC.....	161
Table 7.1 The frequency of collocational errors of dimensional adjectives in the YELC....	167
Table 7.2 Examples of collocational errors of dimensional adjectives (deviant associations)	168
Table 7.3 The error occurrence rates of collocational errors of dimensional adjectives according to error source	169
Table 7.4 The five most significant pre-modifying adjectives of <i>accident(s)</i> sorted by <i>t</i> -score	173
Table 7.5 The five most significant pre-modifying adjectives of <i>attention(s)</i> sorted by <i>t</i> -score	174
Table 7.6 The five most significant pre-modifying adjectives of <i>loyalty (loyalties)</i> sorted by <i>t</i> -score	176
Table 8.1 The distribution of the matched and mismatched non-alternating verbs including the cases of Type 2 (BE + base form).....	192
Table 8.2 The distribution of the matched and mismatched non-alternating verbs excluding the cases of Type 2 (BE + base form).....	193
Table 8.3 Error occurrence rates for L1 morphology according to proficiency level.....	196
Table 8.4 Results of chi-square tests between proficiency level groups.....	198
Table 8.5 Error occurrence rates of each individual verb.....	199
Table 8.6 Error occurrence rates for subject animacy.....	203
Table 8.7 Error occurrence rates according to L1 morphology and subject animacy	204
Table 8.8 Error occurrence rates for subject animacy according to proficiency level	206

Table 8.9 Error occurrence rate of each individual verb and the results of chi-square test ..	208
Table 9.1 Error frequency according to different kinds of affixes.....	213
Table 9.2 Error occurrence per millage according to proficiency level.....	217
Table 9.3 Results of chi-square test comparing proficiency level groups.....	218
Table 9.4 Categorisation of derivational morphological errors based on the corpus-based error analysis.....	220
Table 9.5 Error frequency according to error type.....	225
Table 9.6 Error frequency of Type D and the interlingual influence.....	231
Table 10.1 Overview of spelling errors in the YELC retrieved through the corpus-based error analysis.....	234
Table 10.2 Comparison between frequencies of spelling errors and correctly spelled words in the YELC according to the corpus-based error analysis.....	235
Table 10.3 Examples of interlingual spelling errors.....	237
Table 10.4 Examples of intralingual spelling errors.....	238
Table 10.5 Comparison between interlingual spelling errors and overall spelling errors according to proficiency level.....	241
Table 10.6 Results of statistical tests for interlingual influence comparing proficiency level groups.....	243
Table 10.7 Comparison between intralingual spelling errors and overall spelling errors according to proficiency level.....	245
Table 10.8 Results of statistical tests for interlingual influence comparing proficiency level groups.....	247
Table 10.9 Error occurrence rates of subcategories of interlingual spelling errors according to proficiency level.....	248
Table 10.10 Error occurrence rates of subcategories of intralingual spelling errors according to proficiency level.....	250
Table 11.1 The new error taxonomy according to lexical domains and sources.....	255

List of Figures

Figure 1.1 Schematic representation of the study.....	1
Figure 2.1 James's (1998) lexical error taxonomy.....	23
Figure 3.1 Three types of word knowledge in bilinguals.....	53
Figure 3.2 Two hypotheses of cross-language connection.....	55
Figure 3.3 Revised Hierarchical Model.....	56
Figure 3.4 A blueprint for the speaker.....	66
Figure 3.5 The multilingual processing model.....	69
Figure 3.6 The lexical representation of the mental lexicon.....	71
Figure 3.7 The three stages of L2 vocabulary acquisition.....	71
Figure 3.8 The L2 learners' mental lexicon based on the L2 lexical development model ...	82
Figure 3.9 The lexical error production.....	84
Figure 4.1 The classification of English intransitive verbs.....	104
Figure 6.1 The research procedures for analysing derivational morphological errors.....	159
Figure 7.1 The error occurrence rates of collocational errors of dimensional adjectives	170
Figure 7.2 The picture function for <i>accident(s)</i> in the BoE with its collocates in each column sorted by raw frequency.....	172
Figure 7.3 The concordance lines of the collocation <i>small accomplishment(s)</i> in the BoE ..	177
Figure 7.4 Some of the concordance lines of the collocation <i>great difficulty</i> in the BoE	179
Figure 7.5 The concordance lines of the collocation <i>heavy punishment(s)</i> in the BoE.....	180
Figure 7.6 The concordance lines of the collocation <i>high (and low) confidence(s)</i> in the BoE	183
Figure 8.1 Error occurrence rates for L1 morphology by proficiency level.....	197
Figure 8.2 Error occurrence rates of each individual verb.....	200
Figure 8.3 Error occurrence rates for subject animacy by proficiency level.....	207
Figure 9.1 Error occurrence rate according to proficiency level.....	217
Figure 9.2 Error occurrence rates of each error type.....	226
Figure 10.1 Comparison of error occurrence rates between interlingual spelling errors and overall spelling errors according to proficiency level.....	242
Figure 10.2 Comparison of error occurrence rates between intralingual spelling errors and overall spelling errors according to proficiency level.....	246
Figure 10.3 The interlingual error occurrence rates according to proficiency level	249
Figure 10.4 The intralingual error occurrence rates according to proficiency level	251

Conventions Used in this Thesis

‘single quotation marks’ ~ used for emphasis.

“double quotation marks” ~ used for direct quotations from other studies.

Italics ~ used to indicate words or phrases under analysis or discussion.

Other symbols and remarks

- a. * (asterisk) is used to indicate ungrammatical words or structures.
- b. < (angle bracket) is used to indicate the intended words or structures of the ungrammatical ones.
- c. [] (square brackets) are used to indicate ellipsis in citations or specific letters pronounced by phonemes.
- d. [sic] is used to indicate that the specified words or sentences are transcribed exactly as found in the source text, so that they may contain errors.
- e. This thesis follows the Yale Romanisation System when romanising Korean scripts into alphabets.
- f. The extended examples are numbered separately to the KWIC (keyword in context) lines (or concordance lines) within each chapter.

Main Abbreviations Used in this Thesis

Corpora:

BNC	British National Corpus
BoE	Bank of English
ICLE	International Corpus of Learner English
YELC	Yonsei English Learner Corpus

Other abbreviations:

CA	contrastive analysis
CEA	computer-aided error analysis
CIA	contrastive interlanguage analysis
EA	error analysis
EFL	English as a foreign language
ELF	English as a lingua franca
ELT	English language teaching
ESL	English as a second language
FLT	foreign language teaching
LCR	learner corpus research
L1	first language
L2	second language
RQ	research question
SIGIL	statistic inference: a gentle introduction for linguists
SLA	second language (L2) acquisition

List of Frequently Used Terms

Semantic Domain:

Dimensional adjectives are the adjective forms of dimensional expressions that describe semantic representations in the three-dimensional world, e.g. height, length, width and weight.

Figurative expressions are words or expressions that are used in a figurative sense. These include metaphor, metonymy and simile. In this study, the term **figurative extension** refers to any type of figurative expressions that can be regarded as ‘standard English’. **Figurative overextensions** are regarded as deviant expressions.

Syntactic Domain:

Unaccusative verbs are a type of intransitive verbs that adopt a thematic role Theme as their subject in a sentence. **Alternating unaccusative verbs** have transitive counterparts so that they can be passivised, whereas **non-alternating unaccusative verbs** cannot be passivised because they have no transitive counterparts. This study selects six non-alternating verbs and divides them into two kinds: **matched** and **mismatched** verbs. The matched non-alternating unaccusative verbs are the ones whose Korean equivalents behave in the same manner in terms of passivisation, whereas the mismatched verbs are the ones whose Korean equivalents do not behave alike.

Subject animacy effect refers to influence that a subject in a sentence can have on verb passivisation depending on whether that subject is animate or inanimate.

Morphological Domain:

Metalinguistic awareness refers to the ability to reflect on and manipulate different linguistic units. It includes phonological, orthographic and morphological awareness of language forms.

Morphological awareness refers to a learner’s grasp of morphological structure and their ability to use this knowledge. It can be divided into three components: awareness of inflectional, compounding and derivational morphology.

Derivational morphology is a type of word formation. Awareness of derivational morphology has three aspects: relational, syntactic and distributional knowledge. It may help learners to expand their vocabulary.

There are generally two kinds of affixes: prefix and suffix. In this study, prefixes are called **class-maintaining affixes** in that they form new words without altering the grammatical class but add new information. On the other hand, suffixes are called **class-changing affixes** because they always alter the grammatical class.

Phonological/Orthographic Domain:

Orthographic depth refers to the degree of transparency of sound-to-spelling correspondence in a language. English is considered to have a ‘deep’ orthography where there is a more complex sound-to-spelling relation. On the other hand, Korean seems to have a ‘shallow’ orthography.

CHAPTER 1: INTRODUCTION

1.1 Overview of the Thesis

The motivation for this thesis originated in the challenges confronted during my ten-year career as a language teacher in public primary schools in Seoul, Korea. While the majority of school students learn English as a Foreign Language (EFL) and have great enthusiasm for the subject, I have found that they make particular types of errors in the course of L2 acquisition. My belief that analysing learners' errors would benefit English language teaching in South Korea created the impetus for this research.

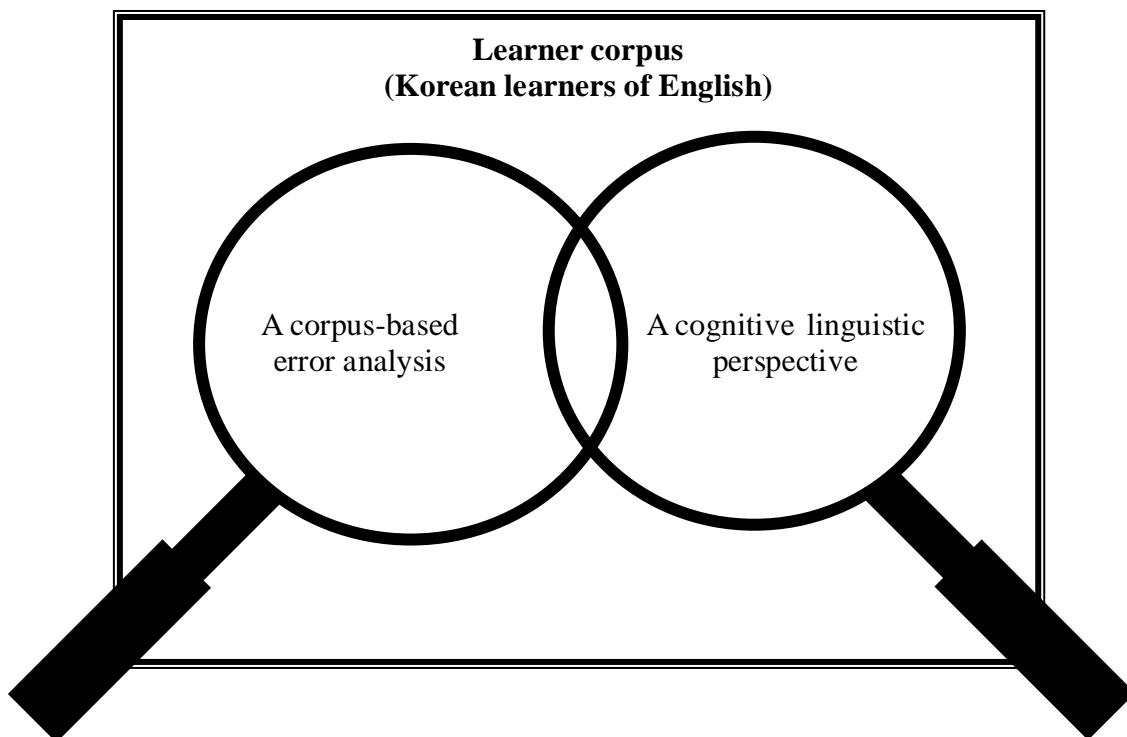


Figure 1.1 Schematic representation of the study

Many previous studies have analysed L2 learners' errors. However, in surveying the existing research I discovered two major weaknesses. First, to the best of my knowledge, almost all previous studies have investigated errors using small data samples, which raises concerns with regard to representativeness (Granger, Gilquin & Meunier, 2015). Secondly, while they have categorised types of errors made by L2 learners, often they have not been able to discuss what kinds of influences or sources affect the error-producing procedures (see Section 2.1.5), which might have offered valuable insights for the English classroom. This was the starting point of the current study.

In order to overcome the weaknesses of previous studies, this research adopts a framework based on three key concepts (see Figure 1.1). The learner corpus, represented by the square, will be viewed by two analysing lenses: a corpus-based error analysis and a cognitive linguistic perspective, respectively the method and the perspective that will be applied in this study.

First, this study is learner corpus research (LCR). A learner corpus (plural: corpora) can be defined as “electronic collections of authentic FL [foreign language]/SL [second language] textual data assembled according to explicit design criteria for a particular SLA [second language acquisition]/FLT [foreign language teaching] purpose” (Sinclair, 1996; as cited in Granger, 2002: 7). In the study of second language acquisition (SLA), for many years researchers have used relatively small samples of mainly artificial data resulting from highly controlled language tasks (Granger et al., 2015). However, learner corpora are key resources, in that written or spoken corpus data produced naturally by L2 learners can provide insights relevant to L2 acquisition and English language teaching (ELT) (see Section 2.3.1).

Second, LCR has often been accompanied by a new type of error analysis (EA) (Granger, 2009), which I will refer to in this study as ‘corpus-based error analysis’. This error analysis method utilises corpus tools that are appropriate for dealing with a large dataset produced by L2 learners (the learner corpus) (see Section 2.3.2). Therefore, the corpus-based error analysis using a learner corpus of essays written in English by Korean undergraduates will contribute to overcoming the first weakness of previous research.

Third, the approach which has been adopted in this research toward L2 learners’ lexical errors has been influenced by models and ideas taken from the field of cognitive linguistics. Cognitive linguistics is a modern school of linguistic thought, characterised by the significant role of cognitive processes in language phenomena in the human mind. Previous research has used diverse ways to classify errors, resulting in a variety of error taxonomies (James, 1998; Llach, 2011). Quite a few studies have tried to speculate as to possible sources. However, even these studies have applied somewhat ‘fuzzy’ error categories (see Section 2.1.1). In contrast, from a cognitive linguistic perspective, the current study presents a clearer categorisation of errors and focuses on L2 learners’ mental lexicon, which comprises different lexical domains in which learners use cognitive processes and are affected by both interlingual and intralingual influences in the production of lexical errors (see Section 3.5). Hence, this perspective will help to provide a relevant account of the second weakness of previous research.

In short, this study aims to analyse a large dataset to detect the errors made by Korean learners of English and to speculate as to the possible sources of those errors. The three key concepts will each be addressed in more detail in Chapter 2. In the following sections, I will briefly address the rationale, objectives and organisation of this study.

1.2 Rationale of the Study

1.2.1 The Significance of Learners' Errors for English Language Teaching (ELT)

Learners' errors are of great interest to researchers in the field of L2 acquisition and ELT, because they offer very important clues as to how L2 learners develop the interlanguage,¹ and provide opportunities to improve teaching materials and strategies in ESL/EFL settings. As Corder (1981: 35) indicated: "The errors that learners make are a major element in the feedback system of the process we call language teaching and learning. It is on the basis of the information the teacher gets from errors that he varies his teaching procedures and materials, the pace of the process, and the amount of practice which he plans at any moment."

It is in that context that error analysis (EA) emerged and became widespread in the field of language research in the 1970s. EA can be defined as "the process of determining the incidence, nature, causes and consequences of unsuccessful language" (James, 1998: 1). Many researchers have investigated learners' errors in the belief that such analysis will help teachers and students to understand learners' weaknesses and find ways to improve them. During the 1980s, interest in EA began to decline, presumably because of the limitations inherent in the approach (see Section 2.1.1 for more information about EA). However, since the end of the 1990s, EA has been reinvented in the form of computer-aided error analysis (CEA), a type of learner corpus research that uses the latest computer technology in combination with a large collection of learners' interlanguage (Dagneaux, Dennes & Granger, 1998).

¹The term 'interlanguage', coined by Selinker (1972), refers to the linguistic system that second language learners have. It may explain the set of utterances produced by second language learners, which is not identical to that of native speakers and reflects learners' attempted production of a target language norm.

One very important consideration in EA is how to categorise learners' errors, a question linked inextricably with the choice of criteria applied to classify errors and finally to establish an error taxonomy. As briefly mentioned in Section 1.1, researchers have developed their own error taxonomies according to various error criteria. The fact that these criteria were sometimes ambiguous might account for one of the major limitations of traditional EA (see Section 2.1.1). The error categorisation is very important in EA studies, because how to define and classify errors is closely related to how these would be used to inform pedagogic practice.

It is for this reason that the current study uses models and ideas taken from cognitive linguistics to categorise L2 learners' errors. This research places particular emphasis on L2 learners' lexical errors and their possible sources, as shown in the L2 lexical development model used as a framework to conceptualise the L2 mental lexicon (see Section 3.5.1). As Carrió & Mestre (2014: 99) indicated:

Error analysis has helped in the understanding of error not merely as an unwanted phenomenon in language, but as a source of information which can be used to improve production in a second language. The errors found in writing can illuminate the writing process and help us to understand the mechanisms that the non-native speaker adopts. As a result, by understanding these error patterns, several strategies may be designed to improve writing in a second language and several different issues such as the cognitive process of language production could be considered when analysing errors.

Carrió & Mestre, 2014: 99

In view of this, the current study focuses particularly on the possible sources of the errors. By doing so, it aims to make meaningful contributions to raising awareness of specific lexical errors and to developing appropriate remedial instructions for L2 learners.

1.2.2 Pedagogic Value of the Study

As mentioned, this study adopts a corpus-based method to analyse a large dataset, namely a learner corpus consisting of essays written by Korean learners of English (see Sections 2.3.2 and 6.2.1). Learner corpora can provide data that are authentic, productive and very large, and thus represent a significant development for L2 acquisition research, which has traditionally been “based on a relatively narrow empirical base, focusing on the language of a very limited number of subjects, which consequently raises questions about the generalizability of the results” (Granger, 2002: 6).

Meunier (2002) shows how LCR can benefit EFL in three aspects: classroom EFL grammar teaching; curriculum design; the production of reference tools. All of these are closely related to raising awareness of L2 learners’ errors among three main stakeholders of language teaching and learning: learners, teachers and researchers. Classroom EFL grammar teaching involves learners and teachers. Curriculum design is the concern of curriculum designers and the authors of textbooks. The production of reference tools may interest lexicographers or language researchers as well as teachers and learners. As such, the findings of the current study will provide valuable pedagogic implications for all of these stakeholders.

Some previous studies have attempted to investigate textbooks in comparison to learners’ interlanguage or a reference corpus, in order to find out how to improve teaching materials (e.g. Römer, 2004; Tono, 2002, 2004). Similarly, the findings of this thesis can be preliminary sources for future research that could show, for example, whether or not Korean EFL textbooks have adequately addressed the four features (see Section 3.5.2) discussed in this study with regard to the interlingual and intralingual influences in the L2 acquisition. Moreover, the findings of this study can be employed to inform the development of student-

centred materials. In addition, since the current study also shows error occurrence rates of error features according to proficiency level (low, intermediate and advanced levels), teaching materials or English curricula can be designed to take these findings into account and to attend to typical error patterns in a larger picture.

In summary, this LCR study is intended to provide a new perspective for language teachers on how to teach English; for L2 learners on how to improve their English proficiency; and for English textbook writers or curriculum designers (or researchers) on how to improve teaching materials.

To reiterate, unlike previous EA studies, the current study can be characterised as follows:

- 1) The approach to analyse the lexical errors in the current study is influenced by models and ideas from cognitive linguistics. That is, this study attempts to identify and describe the sources of the errors based on how learners' mental processes are involved in the production of the errors. Consequently, the current study could provide meaningful and in-depth explanations of the error sources (see Section 2.2).
- 2) Based on the approach as in 1), the lexical errors are classified in its own manner. A new error taxonomy is proposed based on the L2 lexical development model, where the error sources (L1 or L2) and the four lexical domains in the L2 mental lexicon constitute the two criteria used to establish a new error categorisation (see Section 3.2.3). L2 learners' lexical representations can be affected by L1, an effect known as interlingual influence, which results in interlingual errors. On the other hand, L2 learners could also be affected by the target language (L2), known as intralingual influence, which results in intralingual errors (see Section 3.5.2 for the proposed new taxonomy).
- 3) Corpus linguistic tools are employed to analyse a learner corpus using both quantitative and qualitative methods. These methods make it possible to gain objective findings on Korean learners' interlanguage, whereas previous research has tended to rely on researchers' intuition to detect different kinds of L2 learners' errors (see Section 2.3).

1.3 Objectives of the Study

This study aims to answer two main research questions:

- 1) How do interlingual and intralingual influences affect the production of L2 learners' lexical errors?
- 2) Is it possible to categorise the lexical errors according to their sources and domains as presented in the proposed new error taxonomy?

In order to address the first research question this study adopts a corpus-based error analysis, which is able to deal with a large dataset in order to identify the interlingual and intralingual influences that contribute to learners' lexical errors. The dataset investigated by this research is the Yonsei English Learner Corpus (YELC) (see Section 6.2.1), which consists of writings produced by Korean university students. I conducted a preliminary examination of randomly chosen data samples from the YELC (see Section 6.3) and found that Korean learners frequently make errors with four features from the four domains in the L2 mental lexicon, based on the proposed L2 lexical development model. More specifically, I identified collocational errors of dimensional adjectives in the semantic domain, over-passivisation errors of non-alternating unaccusative verbs in the syntactic domain, derivational morphological errors in the morphological domain and spelling errors in the phonological/orthographic domain (see Table 3.1 in Section 3.5.2). Hence, these four error features are selected to be thoroughly investigated in the study.

With regard to the second research question, this study investigates L2 learners' mental lexicon in an attempt to find the sources of the lexical errors. Although many previous studies have analysed L2 learners' errors, almost all of them have focused on how to systematically categorise the errors in a somewhat superficial manner (see Section 2.1). As a result, they

have tended to neglect the sources or factors of the lexical errors that would be expected to give valuable insights and practical implications for the field of ELT. In contrast, in an attempt to investigate the possible sources of the errors, the current study proposes an L2 production model based on the L2 mental lexicon, i.e. the L2 lexical development model (see Section 3.5.1). The lexical errors can be categorised based on the proposed model according to the sources and domains.

1.4 Organisation of the Thesis

The following chapters of this thesis aim to contextualise the current research. They provide the theoretical background to and the methodology employed in the study, discuss the findings regarding the four error features, highlight the pedagogic implications, acknowledge the limitations of the study and offer suggestions for future research.

Chapters 2 to 5 provide the theoretical bases for the study. In Chapter 2, I review the theoretical and historical backgrounds of three key areas that are closely related to the three key concepts of this research, as presented in Figure 1.1 (see Section 1.1): EA; cognitive linguistics; LCR. Chapter 3 provides an overview of the mental lexicon as it has been discussed in previous studies and presents three language production models, based upon which I propose the L2 lexical development model as a conceptual framework for the study. In addition, the chapter proposes a new error taxonomy, which comprises the four lexical domains in the L2 mental lexicon and the sources of errors (interlingual and intralingual influences). Chapters 4 and 5 review the literature for each of the error features investigated in this study, one from each of the four lexical domains.

The methodological basis of the study is explained and discussed in Chapter 6. I set out the research questions (and sub-questions) to be addressed in order to achieve the objectives of the study; describe the learner corpus and reference corpora as well as corpus software programs and statistical measurement tools used in the study; and explain the preliminary examination and the research procedures for each error feature investigated, with reference to specific target items.

Chapters 7 to 10 report the results for the four error features investigated in this research: collocational errors of dimensional adjectives in the semantic domain (Chapter 7); over-passivisation errors of non-alternating unaccusative verbs in the syntactic domain (Chapter 8); derivational morphological errors in the morphological domain (Chapter 9); spelling errors in the phonological/orthographic domain (Chapter 10).

Chapter 11 concludes the thesis by revisiting the proposed new error taxonomy. After summarising the main findings, I discuss the strengths of the study and point out its limitations. Based on these contributions and limitations, some suggestions are made for future research. Finally, I discuss the implications and impacts of the study.

CHAPTER 2: ERROR ANALYSIS, COGNITIVE LINGUISTIC PERSPECTIVE TO L2 LEARNERS' ERRORS AND LEARNER CORPUS RESEARCH

This study adopts a ‘corpus-based error analysis’, in which L2 learners’ errors detected in a learner corpus are discussed according to a cognitive linguistic approach. Therefore, this chapter reviews the theoretical background of the three key areas that relate closely to the current study: error analysis (EA), cognitive linguistics and learner corpus research (LCR).

In Section 2.1, I cover the theoretical bases of EA and previous studies of L2 learners’ errors, especially lexical errors, as well as the limitation of previous error taxonomies, namely the fuzzy criteria (Dagneaux et al., 1998). In Section 2.2, I present the theoretical bases of cognitive linguistics and the significance of using ideas from cognitive linguistics in L2 acquisition and lexical errors. In Section 2.3, I briefly address corpus linguistics and learner corpus research (LCR) and introduce two kinds of corpus-based methodologies applied in the current study. I also briefly review LCR carried out over the last twenty years with specific reference to Korean learners of English.

2.1 Theoretical Background of Error Analysis

2.1.1 Error Analysis and Language Teaching and Learning

Learners’ errors have traditionally been considered as very important resources in the field of second language (L2) acquisition, language teaching and learning (James, 1998). They have been viewed as windows into learners’ interlanguage, through which teachers and researchers

can gain insight as to which aspects of the language pose a learning burden for students. From the 1960s, when S. Pit Corder argued the significance of learners' errors, EA became widespread and the so-called 'EA movement' (Dulay, Burt & Krashen, 1982) saw its heyday in the 1970s. Indeed, EA has helped teachers make major strides toward improving learners' proficiency.

As Corder (1981: 1) claims: "There have always been two justifications proposed for the study of learners' errors: the pedagogical justification [...] and the theoretical justification." According to Corder (1981), the pedagogical justification refers to the contribution of error analysis to language teaching and learning, whereas the theoretical justification regards the understanding of L2 acquisition. From a pedagogic perspective, learners' errors can be significant in three different ways (Corder, 1967): they tell the teacher what the learner needs to learn; they give the researcher evidence of how language is acquired; and (most importantly) they provide learners with opportunities to test their hypotheses about the nature of the L2.

However, EA has been criticised as subject to a number of major weaknesses. Dagneaux et al. (1998: 164) summarised the following five limitations of EA:

- 1) EA is based on heterogeneous learner data.
- 2) EA categories are fuzzy.
- 3) EA cannot cater for phenomena such as avoidance.
- 4) EA is restricted to what the learner cannot do.
- 5) EA gives a static picture of L2 learning.

According to Dagneaux et al., limitations 1) and 2) are methodological weaknesses. The first seems to result from researchers' difficulty in obtaining sufficient relevant data to be

generalisable. Most EA researchers in the 1960s and 1970s tended to examine small datasets, according to what they were able to collect in the given circumstances (e.g. Dušková, 1969; Laufer, 1988; Zimmerman, 1987). This inevitably led to a certain degree of subjectivity and arbitrariness in the data (Da Rocha, 1980). As Meara (1984: 229) argued: “Clearly [...] though these error analyses may provide us with some useful preliminary data, they do not on the whole take us very far.” He seemed to believe that error taxonomies have little predictive or explanatory power because they are essentially post hoc analyses. In other words, the taxonomies are designed after the errors have been analysed, so that it is very difficult to apply a certain error taxonomy to other, different, data (Llach, 2011).

The second limitation listed above has also been pointed out by many other researchers (Ellis, 1994; Llach, 2011). According to Dagneaux et al. (1998: 164), “they [error categories] are often ill-defined, are based on hybrid criteria and involve a high degree of subjectivity.” These unreliable categories of errors could be attributed not only to the absence of clear criteria to categorise learners’ various errors, but also to the fact that errors per se can be viewed in different ways from different perspectives (see Section 2.1.2).

The other three limitations are related to the scope of EA. Since EA focuses on learners’ overt errors, which are the essential part of EA, these limitations seem to be very natural and inherent weaknesses and may be difficult to overcome. Nonetheless, researchers have tried to address them in an attempt to overcome or complement them. For example, regarding the fourth limitation, one could employ complementary measures (e.g. ‘performance analysis’ in Færch’s (1978) study, as cited in James, 1998), where one might compare what learners ‘can’ do with what they cannot do, by investigating instances of correct use as well as those of incorrect use. However, this would not solve completely the third limitation regarding

‘avoidance’, if learners do not produce the target vocabulary or expressions that the EA is dealing with.

In spite of these limitations, researchers continued to conduct EA research, because knowing language learners’ errors is the first step to discovering how to help those learners (James, 1998). Furthermore, the fast-growing field of computer learner corpus research has shed light on a new direction in EA studies (Dagneaux et al., 1998). A corpus-based approach to learners’ errors enables researchers to deal with tremendously large datasets more efficiently and objectively, and this might compensate for the major limitations of EA so that the newly emerging computer-aided EA analysis can contribute to L2 acquisition and language teaching and learning (Granger, 2009) (see Section 2.3).

2.1.2 Error Categorisation Procedures

An EA procedure prototypically consists of four major steps: error detection, error location, error description and error categorisation (also known as error classification) (James, 1998). Although there is a final (fifth) step, i.e. error correction, which concerns how to offer feedback and remedial instructions to language learners, it might be considered an additional or supplementary part of EA because it deals with the implications of knowledge that EA investigates.

The four major steps can be briefly described as follows: when one can identify or ‘detect’ learners’ errors, then one should ‘spot’ the errors in the right location. The error location is discovered when the researcher points out the places in which the errors occur. The location of errors is not always straightforward, because sometimes it might be difficult to find out the learner’s intention. For example, if a learner makes an error in subject-verb agreement, then

in order to spot the error location one should figure out whether the error occurs in the subject or in the verb. Usually, one should be able to spot the place from its context, although this is not always the case. After determining the error location, one should describe the characteristics of the given error. It will then be possible to classify the type of errors according to these descriptions. In other words, the main purpose of EA is to categorise errors, which makes it possible to reveal which errors are the same and which are different. Consequently, the first three steps (error detection, location and description) in an EA procedure are the necessary elements for the fourth step, putting the errors into categories.

In principle, the criteria should be based on precisely observable facts of error descriptions; however, they are not always so clear-cut. This is one of the reasons why EA has been criticised, as mentioned in Section 2.1.1. Another factor that makes error classification more complex is that there could be many kinds of criteria through which learners' errors are viewed. This derives from the nature of errors that learners make. For example, a Korean learner produced an incorrect adjective error *undirect*, whose target form is *indirect*. From an orthographic perspective, it could be viewed simply as a spelling error and classified under such a category. However, it could also be seen as a morphological error, if one perceives this form as an incorrect combination of an adjective *direct* and a prefix *un-*. Again, this error could also be viewed as an L1 transfer error, if one intends to classify the errors according to the factors that cause them (see Section 3.4). These are just some of the possible criteria to classify the adjective error *undirect*, and the example clearly shows that the range of criteria is not confined to linguistic features, but can be extended to, for instance, semantic factors or even sociolinguistic factors such as learners' sex, age, nationality, or education. Once the errors have been consistently categorised, the list is usually called an error taxonomy. However, while the term 'taxonomy' is widely used to mean collections of the classified

errors, a taxonomy is not a mere list of errors, but must be organised “according to certain constitutive criteria” (James, 1998: 102).

In summary, the categorisation of learners’ errors is an essential part of EA, and is preceded by three preliminary procedures: error detection, location and description. Some researchers have tried to establish error taxonomies using their own perspectives and these taxonomies can probably be considered the most important parts of EA procedures.

2.1.3 Historical Background of EA Studies: Two Phases

As mentioned in Section 2.1.1, following Corder’s (1967) introduction of the significance of learners’ errors, EA saw its heyday in the 1970s and EA studies prospered until the 1980s. Subsequently, probably due to the weaknesses researchers found in EA (see Section 2.1.1), EA studies tended to be neglected or even abandoned by researchers in the field of L2 acquisition (James, 1998).

Most of the EA studies conducted during the era of the so-called EA movement focused on grammatical features of errors rather than lexical features. For example, as Levenston (1979: 147) argued, “the study of second language lexical acquisition has been languishing in neglect”. This ‘neglect’ might be partly because applied linguists tend to implement research related to contemporary linguistic theories in language teaching, and around the 1970s they had less interest in lexis in learner langage (Llach, 2011; Meara, 1984). As Meara (1984: 225) pointed out: “Interlanguage theory has traditionally had very little to say about lexical behaviour of non-native speakers.” Another reason could be that errors involving lexical items were considered to be irregular or idiosyncratic, whereas errors of grammatical features were traditionally regarded as more systematic and thus easier to generalise (James, 1998).

However, since the 1980s lexis has begun to receive more attention alongside the belief that vocabulary is very important in language learning, which has coincided with the decline of EA studies (Llach, 2011). Researchers have also started to realize that the borderline between lexical and grammatical errors is not clear-cut (Hemchua & Schmitt, 2006). Consequently, they have started to focus more on ‘lexis’ rather than on ‘grammar’ in the field of language teaching and learning, and accordingly, a new perspective from which to recognise learners’ errors has emerged, namely lexical error.

In summary, the trend of EA studies from the 1970s to recent years could broadly be divided into two phases: 1) EA studies during the EA movement (from the 1970s to around the 1980s); 2) EA studies after the EA movement (since around the 1980s). The next two subsections will briefly review previous EA research and categorisations of errors during the first and second phases respectively.

2.1.4 The First Phase: Review of EA Studies during the EA Movement

The paradigm of EA in the field of applied linguistics replaced the Contrastive Analysis (CA) that was favoured in the 1950s and 1960s. CA was deeply rooted in the behaviouristic and structuralist approaches, where researchers attempted to find the mismatches between L1 and L2 in order to ‘predict’ learners’ difficulties in the L2 acquisition (Brown, 2007). Under the paradigm of CA, researchers viewed L2 learners’ errors merely as deviations to be avoided and not as something that reflects L2 learners’ underlying language processes. By the late 1960s, researchers began to suspect the reliability of CA, as the predictions of L2 learning difficulty on the basis of CA often turned out to be uninformative or inaccurate (James, 1998).

In contrast, EA assumes that L2 learning is a process of somewhat creative construction in which L2 learners consciously test their hypotheses about the target language (Brown, 2007). Therefore, the initial studies within the EA paradigm postulated this incomplete language system of L2 learners and tried to describe their provisional version of the target language. Corder (1967) introduced the concept of ‘transitional competence’ and later developed this into ‘idiosyncratic dialect’ (Corder, 1981). Nemser (1971: 116) referred to a similar phenomenon, which he termed an ‘approximative system’: “the deviant linguistic system actually employed by the learner attempting to utilize the target language”. Nemser also explained that this system could vary according to proficiency level, learning experience, communication function, personal learning characteristics, and so forth. Among the terms coined to refer to similar systems of L2 learners, probably the best known is ‘interlanguage’, which refers to “a separate linguistic system based on the observable output which results from a learner’s attempted production of a TL [target language] norm” (Selinker, 1972: 214).

Based on the concept of learners’ unique L2 systems, a number of EA studies investigated specific errors in order to provide pedagogical implications to ELT and build their own error taxonomies. For example, George (1972) presented ‘common errors in language learning’, extracted from a broad survey with German learners of English. Similarly, Crewe (1977; as cited in James, 1998) investigated errors of Singaporean learners of English that might be acceptable in Singapore but not in an international setting. Richards (1974) observed several types of errors of L2 learners and categorised these into three different types: interlingual; intralingual; developmental errors. Burt & Kiparsky (1972) particularly focused on syntactic errors made by children; they coined the term ‘goof’ to refer to the syntactic deviation from native adult speakers’ English. They also provided various teaching resources for language

classrooms. Many similar types of studies were conducted, particularly in the 1970s and 1980s.

In order to gain an overview of EA studies and the most commonly used error categorisation during the era of the EA movement, I would like to present a comprehensive analysis of Dulay et al. (1982). They reviewed a number of EA studies that were carried out in the 1980s and summarised them by four error criteria:

- 1) Linguistic category
- 2) Surface strategy
- 3) Comparative analysis
- 4) Communicative effect

Firstly, the linguistic category, possibly the most prevalent criterion to classify error types, is based simply on any language component or particular linguistic constituents. Language components include, for example, phonology, syntax, morphology and semantics, whereas linguistic constituents are the elements of each language component. For example, one of the most popular taxonomies might be based on parts of speech from a grammatical perspective.

Secondly, the surface strategy taxonomy is concerned with how surface structures are changed. For example, L2 learners could ‘omit’ necessary lexical items or ‘add’ unnecessary ones. They could also ‘misform’ or ‘misorder’ them. James (1998: 106) suggested that this classification should be termed “target modification taxonomy”, because “it is based on the ways in which the learner’s erroneous version is different from the presumed target version”.

The prototypical categories of this taxonomy include *omission*, *addition*, *misformation* and

*misordering.*² Each category also consists of a few subcategories, e.g. the category *addition* contains *simple addition*, *double markings* and *regularisation*.

Thirdly, the error categories in a comparative taxonomy are based on the comparisons between L2 errors and learners' mother tongue (L1). Researchers have presented various main categories in this taxonomy, for example interlingual errors, intralingual errors and developmental errors (Richards, 1974). However, again the categories differ according to the classification used by each individual researcher. The process used to develop the comparative taxonomy is similar to that of contrastive analysis (CA), and this taxonomy is also called the diagnosis-based taxonomy (James, 1998), because the categories are based on the factors or sources involved in error production. For example, James (1998) presented four major factors that affect the errors made by L2 learners: interlingual, intralingual, communication-strategy and induced factors. He argued that interlingual errors are produced due to negative L1 transfer (or interference), while intralingual errors are related to learning strategies such as incomplete rule application or overgeneralisation. Communication strategy-based errors are produced when L2 learners try to create target expressions in an indirect or circumlocutory way. Finally, induced errors are those affected by incorrect sources such as teaching materials or teacher talk. Although the factors James (1998) presented do not seem to be clear-cut, but overlap across categories, it is worthwhile attempting to speculate as to the causes of errors. The current study adopts a way to categorise errors similar to that of the comparative taxonomy, in that it focuses on the error sources involved in the error production, although the subcategories are simplified in order to avoid the overlapping areas between the subcategories as presented by James (1998) (see Section 3.5.2).

² James (1998) added a fifth category of his own, i.e. *blends*, since there are some situations in which not just one but two well-defined targets could be involved in the production of such errors.

Fourthly, the communicative effect taxonomy focuses on the degree of communicative effectiveness of L2 utterances that contain errors. Determining error types depends solely on listeners' or readers' comprehension, but not on aspects of the errors themselves. In other words, even though a particular error may appear, on the surface, to be critical, it might not severely hinder the communication between interlocutors. Thus, certain types of errors can be considered serious or simple, depending on the understanding of the interlocutors. The two major subcategories of this taxonomy are: global errors that affect overall sentence organisation significantly and hinder communication; and local errors that affect single elements in a sentence and do not significantly hinder communication.

In summary, during the period of the EA movement, researchers have initially emphasised on the interim system of L2 learners such as 'interlanguage', then they have also investigated specific error types that L2 learners seemed to commonly produce in an attempt to draw pedagogical implications in language classroom. Based on Dulay et al.'s (1982) comprehensive analysis of previous error studies until around the 1980s, which described the characteristics of the four kinds of criteria used to establish the taxonomies, each criterion seems to have its own usefulness, although they seemed to neglect L2 learners' use of lexis.

2.1.5 The Second Phase: Lexical Error and its Taxonomies

The second period of EA research can be characterised by its focus on 'lexis' in the L2 acquisition. This does not necessarily mean that researchers in the first period excluded or ignored lexical aspects of errors. Rather, the interest in lexical errors started to increase in the 1980s, and researchers came to interpret the errors through their own perspectives. In fact, the term 'lexical error' seems to be used ambiguously in the literature. It has been broadly used

to refer to any type of error or deviation from the L2 norm with regard to the use of lexical items. However, because many studies have used this term without a clear or unambiguous definition, the lexical errors have been regarded as either a “ragbag category consisting of all errors that are not grammatically fit” or a “superordinate term that serves as a heading for several other classes of errors” (Llach, 2011: 73).

As an example of a ragbag category, Dušková (1969: 24) reported the orthographic errors of lexis produced by Czech adult learners of English under the category ‘lexical error’. This was probably due to the confusion of words that are orthographically similar, such as *respect-aspect* and *plan-plane*. She also reported another kind of lexical error made by Czech learners of English: semantic errors involving confusion to do with verbs such as *do-make*, *attend-visit*. As such, although there have been quite a few studies on lexical errors, it seems that these studies do not provide an explicit definition of the lexical errors or any explanation of how to detect them, or if any, a very superficial and limited description (see Zimmermann, 1987; Hemchua & Schmitt, 2006; Meara & English, 1987).

On the other hand, some studies have used lexical error as an umbrella term in an attempt to establish their own lexical error taxonomies. For example, one representative lexical error taxonomy, that of James (1998; as cited in Llach, 2011: 78), is shown in Figure 2.1. His taxonomy included not only errors of form, such as misselection, misformation, borrowing and coinage; but also errors of meaning, i.e. confusion of sense relation and collocation error. Although James’s taxonomy did not include all types of lexical errors, it did cover the particular lexical error types that the researcher paid attention to. Likewise, other researchers have developed lexical error taxonomies of their own, depending on the particular error features they wished to emphasise (see Channell, 2013; Dušková, 1969; Laufer, 1988;

Lennon, 1991a; Ringbom, 2001; Salem, 2007; Zimmermann, 1987; Zughoul, 1991; Shalaby, Yahya & El-Komi, 2009; Neff, Ballesteros, Dafouz, Martínez & Rica, 2007).

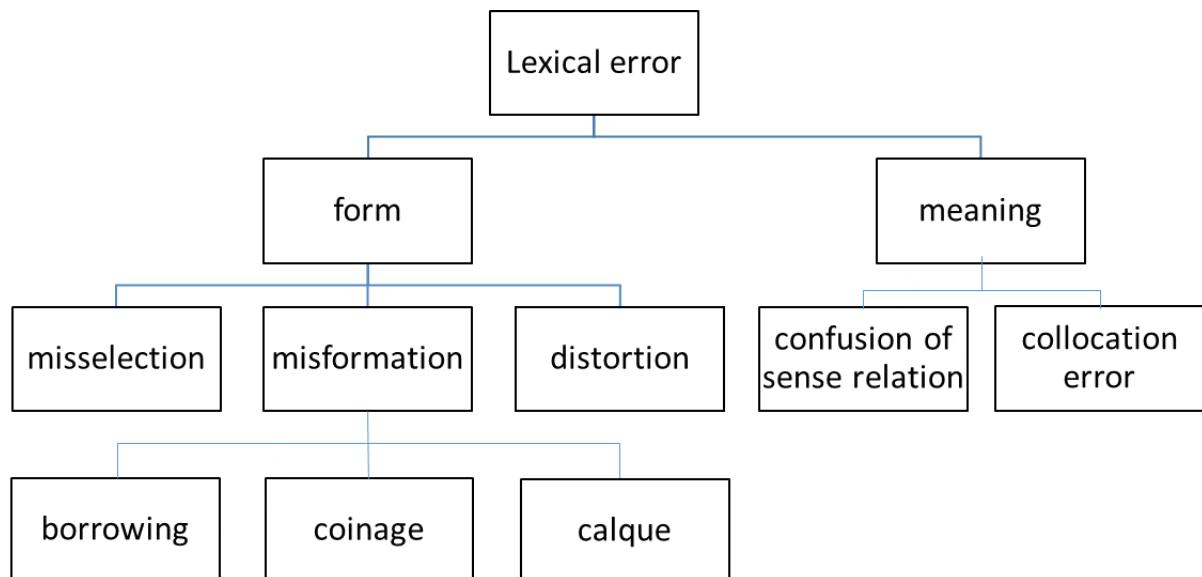


Figure 2.1 James' (1998) lexical error taxonomy³ (cited in Llach, 2011: 78)

Consequently, it was very common across studies that a specific lexical error category in a taxonomy presented by one study could be found in another study in a different sub-category, or even not be included at all. In other words, although a lexical error taxonomy can be established under a specific criterion, some error categories in the taxonomy might also be found in other studies, but under different sub-categories. For example, Ander & Yildirim (2010) investigated compositions written by elementary level Turkish university students (EFL learners) and classified lexical errors into seven different categories. One of these was

³ Llach (2011: 125) disagrees with the position of calque as a formal type of lexical error, because a literal translation (calque) from L1 word to L2 word implies the transfer of semantic features of the L1 word. Here I am in agreement with Llach. Likewise, collocation error could be subsumed within formal error, depending on which aspects of collocation errors a researcher looks at.

‘errors of literal translation’, which could also be regarded in terms of other categories, such as ‘process-oriented lexical error’ or ‘redundancy’, as found in other studies. Llach (2011: 91) also indicated that: “All taxonomies, even the most exhaustive and comprehensive ones, present some problems of overlapping of categories. This implies that there is a changing percentage of lexical errors that could be classified into different types depending on the perspective taken by the classifier as to its source, its cause, and its surface structure or the text.”

Similar to Dulay et al.’s (1982) comprehensive analysis of EA studies carried out during the 1970s and 1980s, as mentioned in Section 2.1.4, I would like to introduce an overview of EA studies after EA movement, particularly about lexical errors. Llach (2011) analysed a considerable body of lexical error studies conducted since the 1970s and divided lexical errors into eight types of taxonomies according to the following criteria:

1) Form- and content-oriented lexical errors:

The distinction between form-oriented and content-oriented lexical errors is probably most common among lexical error taxonomies. It is based on the observation of the L2 mental lexicon, which is organised with formal as well as semantic principles (e.g. James, 1998; Zimmermann, 1987; Carrió & Mestre, 2014; Hemchua & Schmitt, 2006).

2) Descriptive criterion:

The taxonomies are built purely by the surface form of lexical errors without any consideration of their causes or sources. The categories include wrong word choice, omission, addition and misordering (e.g. Hyltenstam, 1988; Lennon, 1991a; Zimmermann, 1986).

3) Etiologic or psycholinguistic criterion:

Researchers focus on the mental processes of L2 learners that result in specific kinds of lexical errors such as overgeneralisation, semantic transfer and confusion of related words (e.g. Dušková, 1969; Engber, 1995). The categories in each study could be

quite different depending on the viewpoints of researchers in recognising the lexical errors.

4) Origin of influence criterion:

These kinds of taxonomies seem to originate in a similar motivation to that of the current study, in that they seek to find what kinds of influences or underlying causes affect the lexical errors (e.g. James, 1998; Richards, 1974; Ringbom, 1987, 2007).

5) Grammatical or linguistic criterion:

The taxonomies guided by this criterion are very similar to those of EA studies in the era of the EA movement (see Section 2.1.4). The lexical errors can also include grammatical or linguistic features of the L2 mental lexicon.

6) Word-class criterion:

These taxonomies classify error types by the class of content words, such as noun, verb, adjective or adverb (e.g. Lennon, 1991b).

7) Product- or process-oriented taxonomies:

Lexical errors can also be categorised according to whether researchers emphasise either the processes that affect the errors or the products of the processes (i.e. the errors themselves). The example studies that adopt product-oriented taxonomies are Engber (1995), Hyltenstam (1988), Lennon (1991b) and Zimmermann (1986, 1987). On the other hand, process-oriented studies include Dušková (1969) and Ringbom (1983).

8) Miscellaneous:

These taxonomies employ a combination of criteria to classify lexical errors (e.g. Zughoul, 1991; Meara & English, 1987).

It is not within the scope of the present study to consider in detail each lexical error taxonomy listed; however, note that each certainly has its own academic and practical value. Indeed, this again shows how difficult it is to define lexical errors and classify them.

The current research, which investigates lexical errors made by Korean learners of English, may possibly belong to one or more of the categorisations mentioned above because it shares some characteristics of those criteria. For example, a new error taxonomy proposed by the current study (see Section 3.5.2) may belong to 3) or 4), in that it focuses on L2 learners' mental processes and the possible causes of the errors. However, the current study differs from previous research since it is based on the conceptual framework of the L2 mental lexicon (see Sections 2.1.7 and 3.5.1).

In summary, during the second phase of the history of EA, a number of EA studies highlighted the lexical errors made by learners as well as their own error taxonomies. Due to the complex nature of learners' errors, any single criterion might be insufficient to cover all of them and it remains possible that other types of criteria could be applied, according to how researchers look at the errors. However, the perspectives offered by previous research do seem to make contributions to pedagogical practices.

2.1.6 A Critical Issue of EA: Errors vs. Mistakes

One critical issue that should be considered in any type of EA is the distinction of two different notions: 'errors' and 'mistakes'. All people, not only L2 learners but also native speakers, make mistakes. However, applied linguists have recognised the importance of distinguishing errors from mistakes. While both errors and mistakes are observed as deviations from a standard, simply put, mistakes refer to occasional lapses in performance, whereas errors are idiosyncrasies that are direct manifestations of interlanguage (Brown, 2007; Ellis, 1994, 1997). In other words, mistakes are "slips of the tongue (or pen)" (Corder, 1981: 10), whereas errors are caused by some kind of systematic failure of L2 processing.

Dulay et al. (1982) explained the difference between mistakes and errors using the terms applied by Chomsky (1965): ‘performance’ and ‘competence’. According to them, mistakes are a particular type of error caused by performance factors such as fatigue or inattention, so they are performance errors. On the other hand, errors are systematic deviations that result from language competence, so are called competence errors.

Corder (1967) argued that the scope of EA should be restricted to errors, and that mistakes of L2 learners should be eliminated from the analysis, because mistakes are not significant to the process of language learning. However, the practical problem in an analysis of learner data is that it is very difficult to distinguish errors from mistakes. James (1998: 83) proposed a way to distinguish them: “Mistakes can only be corrected by their agent if their deviance is pointed out to him or her. [...] Errors cannot be self-corrected until further relevant (to that error) input (implicit or explicit) has been provided and converted into intake by the learner.” However, it seems that the problem still remains, as Brown (2007: 258) indicated: “But the learner’s capacity for self-correction is objectively observable only if the learner actually self-corrects, therefore, if no such self-correction occurs, we are still left with no means to identify error vs. mistake.” Ellis (1997) proposed an alternative approach, that of asking learners to try to correct their utterances. However, this again does not seem to be relevant because learner data that researchers analyse are not always appropriate for discovering the self-correction of L2 learners. Strictly speaking, it seems that it is not only difficult, but indeed almost impossible to distinguish errors from mistakes (Mitton, 1996), particularly when the data from which errors are drawn are large corpus data.

Therefore, this study investigates the learner corpus data with the assumption that every mistake can potentially be regarded as an error. I argue that there are two reasons to consider this assumption as acceptable. The first is a practical problem over the analysis of corpus data.

As mentioned, it seems almost impossible to distinguish errors from mistakes in such a large dataset. Even with a small dataset, the process of distinguishing the two would not be less difficult for a researcher. The second reason is related to the lack of clear theoretical distinction between the two notions. As Dulay at el. (1982) noted, mistakes can also be interpreted as a type of error. In this respect, the distinction between mistakes and errors could rely on the degree of deviation. That is to say, every mistake could be interpreted as an ‘error’ unless it is recognised as ‘self-corrected’.

2.1.7 This Study’s Approach to Lexical Errors

The inconsistency of error categorisation among lexical error taxonomies is not new, and brings us back to one of the weaknesses of EA presented in Section 2.1.1, i.e. that EA categories are fuzzy. With regard to the lexical errors, the inconsistency is particularly salient, to the extent that it could be a major obstacle to systematic and reliable research (see Section 2.1.5).

This obstacle seems to derive not only from researchers’ different perspectives to observe the lexical errors, but also from the nature and origin of lexis. It is not easy to describe lexical error, because such a definition would rely on the notion of ‘lexical competence’. According to Llach (2011: 72): “Depending on what is understood of what ‘knowing a word’ implies, and on what counts as a word, the lexical error can be defined.” That is to say, when learners know a word (or lexis), it means that they not only recognise its phonological sound and its orthographic form, but also understand its various meanings and know how it is used in appropriate ways as well as with appropriate collocates (Nation, 2001). As Llach (2011: 5) indicated, “New words and word respects are acquired in an incremental way”. Hence, the

lexical errors can be produced in various forms, depending on how well learners understand the many aspects of each lexical item.

As mentioned in Section 2.1.4, most of the early EA research was largely concerned with grammatical errors, probably because grammatical errors could be more easily described than, for example, semantic errors involving lexis (Meara, 1984), and accordingly only a few studies have investigated semantic errors of L2 learners of English. However, despite the apparent difficulty involved in examining lexical errors, a focus on lexis in EA might prove to be very useful, in that L2 learners' lexical errors encompass not only grammatical features but also semantic and other features of lexis. Moreover, as it is almost impossible to make an exact distinction between lexis and grammar (Tucker, 1999), it could be reasonable to adopt an integrated perspective to recognise L2 learners' lexical errors.

Consequently, the value of lexical errors and taxonomies in previous studies that contribute to L2 acquisition and language teaching and learning should not be underestimated. As Meara (1984) pointed out, the preliminary data that lexical error taxonomies provide is still valuable, because, as Llach (2011: 74) further explained: "All authors who deal with lexical errors, nonetheless, admit systematicity as a definitional characteristic of errors. Lexical errors are not accidental or random, but respond to systematic causes that can be accounted for in the analysis of the language sample." Lexical errors are particularly significant in the field of L2 acquisition, because they can be evidence of observable interlanguage phenomena (Llach, 2011). Laufer (1991) also suggests that lexical errors can be a reliable instrument to investigate the organisation of the L2 mental lexicon, as well as the vocabulary development of L2 learners. In this respect, it is relevant that the current study investigates L2 learners' lexical errors on the cognitive linguistic basis of the L2 mental lexicon (see Section 3.5.1).

Specifically, it attempts to establish a new error taxonomy based on the sources of the lexical errors under the framework of the L2 mental lexicon (see Section 3.5.2).

In addition, it is encouraging that, in recent years, there have been attempts to establish a more systematic and consistent taxonomy with regard to lexical errors. If one were to design a multi-dimensional error taxonomy with several possible criteria, this could solve one of the weaknesses of the lexical error categorisation, that of ‘fuzzy categories’. Certainly, multi-dimensionality in error categorisation might be challenging and demanding for researchers to apply, since it could involve a lot of effort and expense. However, once the multi-faceted information of errors was fully established, it might facilitate the development of a more consistent and systematic error taxonomy to give a clearer picture of L2 learners’ linguistic performance based on EA.

In one such attempt to build learner corpora with multi-faceted error annotations, Díaz-Negrillo (2009) created a new error tagging system called Error-Annotation and Retrieval System (EARS), which is based on a multi-layer error taxonomy. She divided error categories hierarchically into six linguistic levels: spelling, punctuation, word grammar, phrase grammar, clause grammar and lexis. In turn, the categories in each linguistic level are grouped in four layers: unit identification, error scope, error focus and error type. In this way, Díaz-Negrillo tried to capture different types of errors that can be classified by different kinds of error criteria. Although this example is an attempt at developing an exhaustive and comprehensive lexical error taxonomy, the same kind of approach could help researchers to investigate lexical errors and to develop a more sophisticated and refined lexical error taxonomy from their own particular perspectives.

Although the current research does not aim to establish an exhaustive and comprehensive error taxonomy like that of Díaz-Negrillo (2009), it presents a new error taxonomy, which I believe is more sophisticated and systematic than its predecessors, in an attempt to show a clearer picture of lexical errors by focusing on the ‘sources’ of the errors and the ‘cognitive processes’ involved in the L2 mental lexicon (see Section 3.5.2). To do so, this study explores lexical errors in a learner corpus that comprises naturally occurring data produced by Korean learners of English and employs two theoretical concepts, namely the cognitive linguistic perspective to errors and learner corpus research.

Firstly, in order to speculate as to the error sources, this study adopts a cognitive linguistic perspective (see Section 2.2). Based on the cognitive processes in the L2 mental lexicon, which contains L1 as well as L2 sources, this study presents the L2 lexical development model as a conceptual framework in order to propose a new error taxonomy (see Section 3.5.2). Note that although this study attempts to classify L2 learners’ errors according to the sources that cause them, it does not aim to tackle each and every type of lexical error; instead, it focuses on four specific error features that might be salient, one from each of the four lexical domains (see Chapters 4 and 5).

Secondly, this study is learner corpus research; that is, it adopts a corpus-based error analysis as a method to detect different kinds of lexical errors (see Section 2.3). In contrast to previous studies, in which researchers tended to detect and classify errors with their linguistic intuition, this study utilises reference corpora and corpus tools that enable the handling of large datasets with more objectivity and efficiency. It is hoped that the current study will fill the research gap, to provide a clearer picture of the language behaviours of L2 learners.

These two major concepts, the cognitive linguistic perspective to errors and learner corpus research, will be discussed in more detail in the following sections.

2.2 Cognitive Linguistic Perspective to Learners' Errors

2.2.1 Cognitive Linguistic Perspective to Language Acquisition

Cognitive linguistics, which grew out of cognitive psychology in the 1970s, is a modern school of linguistic thought that attempts to explain language phenomena in the human mind by means of cognitive science. This approach is not a specific theory; rather it is regarded as a ‘movement’ or an ‘enterprise’ that has adopted “a common set of guiding principles, assumptions and perspectives which have led to a diverse range of complementary, overlapping (and sometimes competing) theories” (Evans & Green, 2006: 3). It emerged from dissatisfaction with formal approaches such as Generative Grammar and challenges the idea that humans have a ‘language acquisition device’ that is an autonomous tool for language acquisition. Instead, cognitive linguists describe language phenomena with an assumption that language reflects ‘patterns of thought’. Therefore, they believe that “the cognitive processes governing language use and learning are essentially the same as those involved in all types of knowledge processing” (Littlemore, 2009: 1). In other words, the language acquisition process relates to linguistic knowledge of meaning and form, including not only semantic but also syntactic, morphological and phonological structures of language (Croft & Cruse, 2012). Cognitive linguists raise hypotheses about the conceptual system that are “based on observing patterns in the way language is structured and organised” (Evans & Green, 2006: 15).

The idea that language phenomena are closely related to language users' cognitive processes gives rise to several key concepts in cognitive linguistics in relation to language learning and teaching, many of which have been researched extensively, for example categorisation, encyclopaedic knowledge, metaphor, embodiment and polysemy (Littlemore, 2009). These key concepts can also be investigated within the process of L2 acquisition, because the L2 acquisition can be described as 'conceptualisation of target language', which is in line with one of the assumptions of cognitive linguistics.

For example, with a cognitive linguistic perspective, one may be able to examine L2 learners' acquisition of the polysemous nature of lexical items. Polysemy is defined as "the association of two or more related meanings with a single phonological form" (Taylor, 1995: 99). The majority of lexical items, particularly highly frequent words, are polysemous. As Nerlich & Clarke (2003: 5) indicated: "Multiple (word) meanings have been accounted for in dictionaries for at least two centuries." Let us consider the verb *run*. According to the Collins Online English Dictionary,⁴ the verb *run* has more than thirty senses, which denote slightly different but related meanings in various contexts, e.g., 'I was *running* in the Marathon', 'He laughed and *ran* his finger through his hair', 'He *runs* his own office'. With regard to the L2 acquisition of the item *run*, researchers may investigate such questions as how L2 learners recognise the core meaning and the other senses of *run* and how the different meanings are connected; whether and why learners frequently make errors with specific senses; and whether their mother tongue (L1) affects these errors, from the perspective of cognitive linguistics.

One cognitive linguistic approach, Rosch's prototype theory, provides an interesting standpoint to perceive L2 learners' errors in terms of the polysemy of lexical items. The

⁴ It can be accessed at the URL: <http://www.collinsdictionary.com/dictionary/english>.

‘prototype’ refers to a language phenomenon whereby humans tend to categorise common objects and to create some ideal exemplars (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976). Although most words have a polysemous nature, L2 learners tend to focus on a single prototypical meaning of a word and then to unconsciously ignore its ‘peripheral’ meanings. According to the prototype theory, L2 learners may easily acquire the prototypical meaning, as in ‘*running* in the Marathon’, but at the initial stage of language learning they may not yet be able to realise the other peripheral meanings. Then they link the prototypical meaning of the word or phrase to its L1 equivalent. Especially when L2 learners do not acquire those peripheral meanings of the L2 word or phrase that do not perfectly match with the L1 equivalent, they might get confused and make errors with regard to the peripheral meanings. Therefore, it could be argued that these kinds of L2 learners’ errors can be caused by semantic ‘incongruity’⁵ between L1 and L2. These errors suggest that the meaning structures of polysemous words in the L2 mental lexicon seem to be not irregular or unmethodical, but rather, systematic and well-ordered. Awareness of the cognitive structure of word meaning can help language teaching and learning (Csábi, 2004). Although the prototype theory is concerned with meanings of words, which is a semantic domain, it could be suggested that similar kinds of incongruity could affect the other domains, such as syntax, morphology and orthography.

Furthermore, cognitive linguistic approaches can provide useful accounts of the mental lexicon with regard to L2 acquisition. It is believed that humans store ‘words’ in their mental lexicon or ‘mental dictionary’. However, the mental lexicon is quite different from a book dictionary (Aitchison, 2012): its contents are not fixed and it contains far more information

⁵ In this study, the term ‘incongruity’ is used to refer to a state in which there is no equivalence between L1 and L2. This concept is not confined to the semantic domain of lexis, but also extends to other domains of the L2 mental lexicon (see Section 3.5.1). Similarly, the term ‘lexical non-congruence’ refers to a lack of direct translational equivalence between L1 and L2 (Marton, 1977, as cited in Bahns, 1993).

about lexical entries. According to Aitchison (2012), humans do not just retrieve words when needed, but can create new words and meanings from time to time. Moreover, in the case of L2 learners' mental lexicon, it seems that the lexical representations and language processing procedures are even more complicated (see Sections 3.1 and 3.2).

Researchers may be able to investigate the development of the L2 mental lexicon from a cognitive linguistic perspective. For example, Jarvis (2009) expounded a phenomenon called 'lexical transfer', which suggests that L2 learners' knowledge and use of words can be affected by their L1 mental lexicon. He defined 'lexical transfer' as "the influence that a person's knowledge of one language has on that person's recognition, interpretation, processing, storage and production of words in another language" (Jarvis, 2009: 99). He focused on how the lexical representations are conceptualised in the mental lexicon and provided a useful conceptual framework for lexical transfer to be explored from a cognitive linguistic perspective.

Consequently, cognitive linguistics can finally provide useful implications to language classroom. As Littlemore (2009: 3) states: "One of the contributions that cognitive linguistics makes to second language learning and teaching is to suggest ways in which the relationships between grammatical expressions and their original lexical meanings can be made apparent in the language classroom to enhance learning and memorization. This process encourages learners to explore the deeper meanings of grammatical items and to think about why the target language expresses things the way it does." This suggests that the findings of the current study which adopts cognitive linguistic perspective to lexical errors can provide valuable pedagogic implications. Therefore, this study addresses the categorisation of lexical errors by proposing a relevant L2 lexical development model as a conceptual framework to establish a new error taxonomy (see Section 3.5).

2.2.2 Contributions of the Cognitive Linguistic Perspective to this Study

As has been noted above, cognitive linguistics has contributed to second language teaching and learning by exploring L2 learners' mental processes when they exhibit various language learning phenomena. As such, it might contribute to our understanding of why L2 learners make lexical errors. A cognitive linguistic perspective could help to explain the underlying mental processes of L2 learners when they face incongruity between L1 and L2, or when they lack words in their mental lexicon. Presumably, L2 learners are in the course of acquiring an appropriate mapping between L1 and L2, which is closely related to their mental processes. Hence, investigating L2 learners' errors with a cognitive linguistic perspective requires an examination of how their mental processes cause lexical errors, which include not only grammatical and morphological features, but also semantic and orthographic features of L2 acquisition.

A cognitive linguistic perspective could also help to investigate possible causes of the errors. On the surface, it might look as if the lexical errors of L2 learners stem simply from the lack of storage in the L2 mental lexicon. However, I assume that the possible causes of the lexical errors might be attributable to the lexical information (L1 or L2) that L2 learners refer to and/or to their cognitive processes, when they produce L2 lexical representations. If the possible sources of the errors are related to the L1 or L2 lexical information, then the errors could be classified as interlingual or intralingual error respectively (see Section 3.2.3). For example, L2 learners might make lexical errors by selecting inappropriate lexical items and this incorrect choice might be affected by literal translations of L1, resulting in interlingual errors. However, if the incorrect choices were caused by incomplete L2 rule application, they would be classified as intralingual errors. Moreover, the cognitive processes might also play

roles in causing lexical errors, e.g. communication strategy-based errors (James, 1998) (see Section 3.4).

According to Littlemore (2009: 4): “Different languages conventionally construe things differently, and although we may not be consciously aware of it, it is likely that our cognitive systems will, to some extent, have been ‘primed’ by our first language (L1) in ways which might interfere with our learning of subsequent languages.” In that sense, the lexical errors could be caused by the fact that the conventional use of English has not yet been ‘primed’ (Hoey, 2005) in their mental lexicon. In order to find the possible causes of L2 lexical errors, it is necessary to speculate as to L2 processing procedures, which might be understood with the help of a cognitive linguistic perspective.

2.3 Theoretical Background of Learner Corpus Research

2.3.1 Corpus Linguistics and Learner Corpus Research

It was in the early 1960s that the first modern corpus of the English language, the Brown corpus (i.e. the Brown University Standard Corpus of Present-day American English), was built for American English. Since then, with the rapid growth of computer technology from the 1980s, corpus linguistics has become widespread not only in branches of linguistics, but also in other interdisciplinary studies. Most linguists would agree that corpus linguistics is now a research methodology that provides a new empirical approach that can benefit almost every area of linguistic research, including lexicography, language teaching, translation, stylistics, grammar, gender studies, forensic linguistics, and computational linguistics. As Hunston (2002: 1) argued: “It is no exaggeration to say that corpora, and the study of corpora, have revolutionised the study of language, and of the applications of language.”

Corpus-based research is now almost fully fledged in terms of describing many aspects of English, although there is still much work to be done. However, research into non-native English in terms of L2 acquisition and language teaching and learning remains uncharted, even though such fields are where corpus linguistics can “change both the object to be taught and the way it is taught” (Tognini-Bonelli, 2001: 10). With regard to English language teaching (ELT), Hunston (2002) argues that not only can corpora lead language teachers to new descriptions of a language so that what they are teaching can be changed, but also that corpora themselves can contribute to producing language teaching materials and forming the basis for syllabus design or methodology. After all, “teachers can have more objective information about their students’ difficulties on the one hand, and more powerful tools with which to work on their students’ data on the other” (Díaz-Negrillo & Thompson, 2013: 19).

A corpus can be designed for particular purposes, and there is a specific type of corpus for investigating L2 acquisition or language teaching and learning: a learner corpus (Granger, 2002). Since a learner corpus is a collection of naturally occurring data (written or spoken) produced by learners, the appearance of learner corpora shows a new direction for L2 acquisition and language teaching and learning, one that is of growing interest for language education researchers at large (Granger, 2008).

One important issue we should consider in using learner corpus data in language research is the notion of ‘authenticity’: how similar the data are to real-life language use or how natural they are (Mauranen, 2004). Learner data (written or spoken) have always been a very important resource for studies in the various fields of language and linguistics. Ellis & Barkhuizen (2005) have divided the samples of learners’ production data into three different kinds: 1) naturally occurring samples; 2) clinically elicited samples; 3) experimentally elicited samples, according to the degree of control researchers have on obtaining such data.

The three types constitute a continuum, with naturally occurring samples at one end, indicating that the researcher has no control, whereas at the other end the researcher exercises very close control of the data production, as in experimentally elicited samples, and clinically elicited samples are somewhere in between. The notion of authenticity is somewhat problematic because it is a relative concept, and thus it may be quite difficult to draw a fine line between natural and experimental (Mauranen, 2004).

Learner corpus data can be regarded as ‘authentic’ in that they consist of language produced in situationally and interactionally authentic circumstances (Ellis & Barkhuizen, 2005). However, it should always be noted that ‘authenticity’ is not absolute; as Granger (2002: 8) pointed out: “We all know that the foreign language teaching context usually involves some degree of ‘artificiality’ and that learner data is therefore rarely fully natural.”

Therefore, the emergence of learner corpora brought a new perspective to L2 acquisition research. Until then, data samples were usually too small, often collected from relatively few informants, which gave rise to concerns about representativeness. Furthermore, very few studies have employed ‘authentic’ learner data.

One pioneering example of learner corpora is the International Corpus of Learner English (ICLE) (Granger, Dagneaux, Meunier & Paquot, 2009), which consists of argumentative essays written by advanced learners of English as a second language. Nowadays, there are a number of learner corpora available, which contain learner data from various L1 backgrounds. Nevertheless, Leech’s (1998: xvii) observation that “it is time that some balance was restored in the pursuit of SLA [second language acquisition] paradigms of research, with more attention being paid to the data that the language learners produce more or less naturalistically” suggests that it is quite rare to use naturally occurring data in the L2

acquisition field and a majority of studies still rely on data from experimental tasks such as grammaticality judgement tasks (see Section 4.2.2).

The study of learner corpora, known as learner corpus research (LCR), brings a new perspective on L2 acquisition. As Granger et al. (2015: 1) explained: “[LCR is] a new research strand [that] emerged in the late 1980s as an offshoot of corpus linguistics, a field which had shown great potential in investigating a wide range of native language varieties (diachronic, stylistic, regional) but had neglected the non-native varieties.”

One of the main analyses in LCR is based on the comparison between learners’ data and native speakers’ data as a norm (Hunston, 2002). This has enabled traditional EA to enter a new phase of interlanguage research, where it has been ‘reinvented’ in the form of a corpus-based error analysis by means of learner corpora (Dagneaux et al., 1998). It is the advent of large-scale learner corpora that has made possible this kind of LCR (see Granger, 2002; Meunier, 2002; Nesselhauf, 2004).

The current study undertakes contrastive and/or comparative analysis between a Korean learner corpus of English and English reference corpora or a Korean reference corpus. To do so it makes use of LCR methodologies, which will be briefly covered in the next section.

2.3.2 Learner Corpus Research Methodology: CIA and CEA

In LCR, two types of research methods are typically used (Granger, 1998, 2008): Contrastive Interlanguage Analysis (CIA) and Computer-aided Error Analysis (CEA).

CIA involves quantitative and qualitative comparisons of L1 vs. L2 as well as L2 vs. L2 (Granger, 1998, 2008, 2012). Unlike traditional contrastive analysis, which manually

compares different languages, CIA compares varieties of language output with the help of corpus tools. One method employed in CIA is to investigate comparisons between native and non-native data, and here native corpora can be used as a norm in order to reveal non-native features, including errors, from learner corpora (Granger, 2002). In other words, this method aims at “uncovering the characteristics and patterns of use that distinguish learners from native speakers in terms of quantitative as well as qualitative differences” (Callies, 2015: 40). Although researchers should take care to choose a relevant native corpus as a control, fortunately there is available a wide range of reference corpora of sufficient size to be appropriate as a norm. In the current study, the BoE (Bank of English) will be used as the main reference corpus and the BNC (British National Corpus) as a supplementary reference corpus (see Sections 6.2.2 and 6.2.3).

A typical example of CIA is the study of the verb ‘make’ conducted by Altenberg & Granger (2001). They investigated how EFL learners use (overuse or underuse) *make*, one of the high frequency verbs in English, by comparing learner corpora with a corpus comprising data from native speakers. They used the Swedish and French sections of the ICLE as learner data and the Louvain Corpus of Native English Essays (LOCNESS) as a native speaker norm. The study first showed the overall frequency of the verb in the three corpora and found that French learners of English significantly underused the verb *make*, in comparison to Swedish learners of English and native speakers. It then examined the profile of *make* in detail, according to eight major categories of use, including result of creation, delexical uses, causative uses and phrasal uses. For example, Swedish learners of English significantly underused the delexical *make*, whereas they significantly overused the causative *make*. Finally, the study demonstrated that some of the results seemed to be derived from an

interlingual factor. The research provided some pedagogic implications as a way of raising L2 learners' awareness of the high frequency verb *make*.

CEA, the other method frequently used in LCR, also uses quantitative and qualitative methods to focus on errors produced by learners, which researchers then analyse with the help of computer tools (Granger, 2002). CEA is quite different from the traditional EA studies mentioned in Section 2.1.1, in that it is assisted by computers, hence data processing is extremely fast and consistent. Once learner data are computerised, corpus software tools can simply count the frequency of words, retrieve concordance lines and analyse key words. However, the major challenge facing CEA is that it is often very difficult and time-consuming to annotate with relevant tagging that suits a given study in a learner corpus (Granger, 2004). Yet as Granger (2009: 24) highlights: "A thoroughly error-analysed learner corpus is an invaluable resource which can inform most pedagogical tools." In other words, if CEA is conducted with a well-annotated learner corpus, it will be very practical to analyse learner data to find evidence of specific features or useful results that can be applied to L2 acquisition and ELT.

Thewissen's (2013) study using an error-tagged learner corpus is a typical example of CEA. She conducted a kind of quasi-longitudinal analysis in an attempt to show the dynamics of L2 learners' development. The learner data were the French, German and Spanish sections of the ICLE, where each essay was annotated with error tags according to the Louvain Error-Tagging System and divided according to English proficiency levels. Seven types of errors were categorised and the frequency of each error type was counted. The results demonstrated the error developmental patterns of EFL learners according to the proficiency levels. For example, in the error category 'single choice of lexical items', the learner data illustrated a strong developmental pattern, where the number of errors steadily decreased as proficiency

increased. On the other hand, some error categories, such as ‘relative pronouns’ or ‘conjunction subordination’, displayed a weak developmental pattern in which the error frequency at each proficiency level was such that there was no clear indication of progress. Therefore, the study findings implied that each error category has a specific developmental pattern and this has significant pedagogic implications for linguists and language teachers.

The current study adopts a combined CIA and CEA corpus-based method. It could be seen as a CIA study, in that it compares a learner corpus with native reference corpora in order to find learners’ ‘deviant expressions’.⁶ It could also be regarded as a CEA study, because it deals with learners’ errors on specific linguistic features with the help of computer tools. Both methods will be very useful to detect errors from a Korean learner corpus and to establish a new error taxonomy. Therefore, I propose to use the term ‘corpus-based error analysis’ to refer to the methodology used in this research. The corpora used in the study and the procedures for conducting the corpus-based error analysis, will be introduced in more detail in Sections 6.2 and 6.3.

2.3.3 Corpus-based Research on Korean Learners of English in Korea

Since its introduction in the early 2000s, corpus-based research has captured the attention of scholars, particularly in the field of English language teaching in Korea. Indeed, a number of researchers have been interested in corpus linguistics and have attempted to utilise corpus tools in their work. However, despite this growing interest in the use of corpus-based methods, to date there have been very few studies that have involved with CIA and CEA.

⁶ The term ‘deviant expression’ refers to a form or usage that is unacceptable according to the norm (native English) (Nesselhauf, 2005). It includes the lexical error analysed in the current study.

Goh (2011) pointed out the limitations of corpus-based English vocabulary studies in Korea over the last ten years: in almost every case, the studies did not conduct a systematic analysis at the overall vocabulary level of text, and most of them did not go beyond a superficial comparison between the given texts with a wordlist of a reference corpus.

I have searched for corpus-based research analysing learner corpora in Korea conducted since 2011 and found that there have been very few studies with regard to Korean learners of English. Only in recent years have some researchers started to look into specific features or constructions that occur in learner corpora (see Ha, 2016; Lee, 2017; H. Park, 2015, 2016; Koh & Kim, 2014; Park & Choi, 2016; Ryu & Park, 2017), and to date the research examining phraseology or learners' errors in learner corpora is sparse (see Song, 2018; Kim & Yoo, 2015; Lauzon & Song, 2016; Lee & Kim, 2017), as shown in Table 2.1.

Table 2.1 List of recent learner corpus research in Korea

Author(s)	Year	Specific features or topics of the research
Ha, Myung-Jeong	2016	The use of linking adverbials
Lee, Eun-Joo	2017	The development of English personal reference
Park, Hyeson	2015	Comparisons between 'actually' and 'in fact'
Park, Hyeson	2016	The use of predicational adverbs
Koh, Sungwan & Kim, Bonggyu	2014	The use of modal expressions
Park, Keunhyung & Choi, Incheol	2016	The acquisition of English negatives
Ryu, Miryeong & Park, Mae-Ran	2017	A comparative analysis of coordinator 'and'
Song, Sanghoun	2018	An analysis of phraseology
Kim, J. E. & Yoo, I. W.	2015	An analysis of to-infinitive errors
Lauzon, Kory & Song, Sanghoun	2016	The use of a specific word 'hardly'
Lee, Hye-Kyung & Kim, Hyewon-Okh	2017	An analysis of lexical bundles

The reasons for this comparative lack of LCR in Korea are potentially twofold. First, there are not many learner corpora available in Korea. Although some researchers have compiled small-scale learner corpora for their own studies, i.e. DIY (do-it-yourself) corpora (McEnery, Xiao & Tono, 2006), there are only a few large-scale learner corpora that contain Korean learners of English, e.g. the YELC, the NICKLE (Neungyule Interlanguage Corpus of Korean Learners of English) (Kwon, 2008) and the SNU (Seoul National University) corpus (Kwon, 2009). What is worse, the NICKLE and the SNU corpora are not publicly available, and even for researchers it is difficult to gain access to them. The second possible reason is that corpus-based methods with regard to LCR are not yet well-rooted in the field of English language teaching in Korea. In this respect, the current study will be noteworthy and the findings will provide insightful implications for English language classrooms, particularly in Korea.

2.4 Chapter Summary

This chapter has provided three key concepts as the theoretical bases for the current study: error analysis (EA); cognitive linguistic perspective to L2 acquisition; corpus linguistics and LCR.

In Section 2.1, I discussed the historical and theoretical background of EA and its limitations, as well as the error categorisation and its fuzzy criteria. Although EA was prevalent in the 1960s and 1970s, it has certain clear limitations. However, such limitations can be compensated for with a newly invented corpus-based methodology. Also in Section 2.1, I focused particularly on lexical error and its complex nature. Lexical error could have important practical implications, since lexis cannot be separated from grammar, especially with regard to English language teaching.

In Section 2.2, I discussed how the application of a cognitive linguistic perspective to investigate lexical errors might be able to reveal the possible underlying sources or factors of lexical errors and the mental processes in L2 language production. It is very important to speculate as to how and why L2 learners make such lexical errors, because this will provide significant pedagogic implications for language teaching and learning. The current study assumes that there are two major sources of lexical errors, namely L1 and L2, which result in interlingual and intralingual errors respectively. In the next chapter, this thesis will present a conceptual framework and a new error taxonomy to describe and classify lexical errors based on a cognitive linguistic perspective.

In Section 2.3, I reviewed the basic notion of corpus linguistics and LCR in order to provide the methodology that will be applied in this study. It can be argued that application of the corpus-based approach to a large dataset of learner data (learner corpus) will offer more systematic and objective ways to investigate the learner data. Therefore, the current study adopts a combined CIA and CEA version of corpus-based error analysis in order to investigate particular features of lexical errors produced by Korean learners of English. In addition, I briefly reviewed LCR conducted in the last twenty years in Korea. Given that there has so far been very little LCR conducted with regard to Korean learners of English, the current study is possibly the first corpus-based error analysis with a large-scale learner corpus in Korea.

CHAPTER 3: LANGUAGE PRODUCTION MODEL AND A NEW ERROR TAXONOMY

This chapter presents a relevant language production model as a conceptual framework to investigate specific types of lexical errors from the cognitive linguistic perspective on L2 production discussed in the previous chapter. It also proposes a new error taxonomy based on the L2 lexical development model in order to provide a conceptual basis to categorise the errors.

In Section 3.1, I give an overview of the mental lexicon, a complex language system that processes language storage and retrieval. In Section 3.2, I introduce three theoretical issues in relation to the L2 mental lexicon that have been investigated in previous research. These three issues provide the basis for the perspectives adopted toward the L2 mental lexicon in the current research, in order to provide a theoretical basis for a conceptual framework for the study. In Section 3.3, I outline three language production models in an attempt to arrive at an appropriate L2 acquisition model; that is, one that includes relevant criteria to classify the lexical errors considered in this study. In Section 3.4, I begin by reviewing previous studies involving diagnosis-based error taxonomies and discuss their limitations in terms of the arbitrariness of categorisation of the possible causes of L2 learners' errors. Then, I focus on the 'sources' of lexical errors and the role of 'cognitive vehicles' (see Section 3.4.2) in producing L2 lexical representations. In Section 3.5, I introduce the L2 lexical development model and propose a new error taxonomy for the current study based on two main criteria: the lexical domains and the sources of errors.

3.1 Overview of the Mental Lexicon in the Human Mind

The human mind, in general, has been believed to hold a somewhat complex system known as the ‘mental lexicon’, where human beings store words and from which they retrieve those words when needed (Aitchison, 2012). It is very important to conceptualise the mental lexicon in any language processing model, because it is through such a conceptualisation that we can explore how language information is stored and retrieved. We can explain how form and meaning of a word or multi word units are linked in comprehension and production of language. We may also be able to see how information from different language domains, such as phonology, orthography, morphology, semantics and syntax, is combined.

However, it is by no means simple to provide a straightforward explanation of what the mental lexicon is. Because it exists in the human mind, the mental lexicon is not a tangible or concrete thing; rather, it is a complex abstract system in nature. Therefore, it is likely that many of the attempts at modelling the mental lexicon have been “wrong or, at the very least, incomplete” (Jarema & Libben, 2007: 1). In Aitchison’s (2012: 262) term, the mental lexicon can perhaps be characterised as “an evolutionary mish-mash”. Consequently, over the decades different researchers have given various kinds of accounts in terms of, for instance, what components constitute the mental lexicon, how it stores language information, how it uses that stored information to recognise or produce words, and how one might describe the procedures of comprehension and production. Such speculation continues, and to date there is no clear picture of what the mental lexicon might look like or how it operates in the comprehension and production of language (Aitchison, 2012). Furthermore, when considering the L2 mental lexicon with regard to bilingualism, there are additional theoretical issues that make these accounts more complicated, such as ‘Are there one or two mental

lexicons in one's mind?', 'How is the L1 and/or L2 activated?', 'Do both languages share semantic, syntactic, morphological knowledge?' (Schreuder & Weltens, 1993).

Another reason for the difficulty in presenting a particular definition of the mental lexicon is that a proper definition should be based on some consensus on a theory of the cognitive processes with which such linguistic domains as phonology, morphology, syntax and semantics are associated. To date no such consensus has been reached (Jarema & Libben, 2007).

Nevertheless, despite those obstacles to conceptualising the mental lexicon, researchers' endeavour to discover what the L1 and/or L2 mental lexicon looks like has made major contributions to characterising the underlying mechanism, and such attempts at modelling continue (Aitchison, 2012; Barcroft & Sunderman, 2008). According to Murre (2005), there are two main approaches to the formulation of language acquisition models. One is the 'symbolic-deductive' paradigm that originated with Chomsky in the 1960s, where he insisted that humans have an innate mental language system (i.e. language acquisition device) prestructured for language processing. The other is the 'subsymbolic-inductive' paradigm from connectionism, which is rooted in the work carried out by Skinner during the 1950s. With the help of the rapid increase in computer processing speed and the growth of computer storage capacity, connectionism was revived in the 1980s and, since then, a number of 'computational models' of the mental lexicon have been developed.

The early computational models constructed in the 1970s suffered from a number of limitations, particularly with regard to the L2 (or bilingual) mental lexicon.⁷ As Thomas &

⁷ Researchers in the field of bilingualism may prefer to use the term 'bilingual mental lexicon', whereas the current study chooses 'L2 mental lexicon', because it focuses on the acquisition by L2 English 'learners' rather than 'fluent bilinguals'. However, the two terms can be used interchangeably in this study, without a particular difference in meaning.

Van Heuven (2005: 203) indicated: “[T]hey were often little more than verbal description. There was no precision in the specification of how these recognition systems would work.” In other words, since the early models were dependent on a somewhat ‘metaphoric’ explanation of the bilingual mental lexicon, it was not always possible to verify or falsify their hypotheses. On the other hand, the advent of computational modelling made it possible to simulate empirical data and to evaluate the validity of such models. Researchers sometimes even used functional neuroimaging techniques such as positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), which can show the specific regions of the brain that are working in linguistic activities (Abutalebi, Cappa & Perani, 2005).

However, it appears that there is still no clear picture of the L2 mental lexicon and the exploration to reveal the nature of language acquisition continues (Kroll & Tokowicz, 2005). In spite of the contributions that computational models have made, there are still many questions to be answered, and so there remain different perspectives on the L2 mental lexicon. For example, one of the major questions from the middle of the 20th century in the field of bilingualism is whether the bilingual’s two languages are stored in separate memory systems or a shared one (see Section 3.2.1). Contrary to the expectation that neuroscience approaches would be able to provide a clear answer to this question, it remains unresolved, because it is difficult to confirm that a distinct neurological activity in a brain is definitely linked to a certain lexical representation of a certain language (Channell, 2013; Kroll & Tokowicz, 2005). In addition, these models may be somewhat subjective according to the perspectives of the researchers who apply them. As Jarema & Libben (2007: 3) point out: “The views that a researcher holds with respect to questions such as these will have substantial impact on how that research situates the enterprise of mental lexicon research and what he or she considers to be of key importance.”

Therefore, language processing models in bilingualism studies have been built upon particular theoretical assumptions that could be supported by empirical data, while at the same time there could exist counterexamples from different perspectives. As Aitchison (2012: 263) noted: “The mental lexicon, therefore, seems to be a mixed system which has found a workable compromise between the requirements of production and those of comprehension.”

However, formulating an optimal language processing model is not within the scope of the current research, and therefore this study will inevitably be eclectic in terms of its perspective on the L2 mental lexicon. Accordingly, it should be acknowledged that the current study may to some extent be based on assumptions that, one hopes, will be verified in the near future. The next section will briefly address some of the key issues raised in the field that will help to lay the groundwork to establish the L2 lexical developmental model for the current study (see Section 3.5.1).

3.2 Theoretical Issues in the Models of the L2 Mental Lexicon

Previous studies in the field of bilingualism have addressed a number of theoretical issues in the process of formulating models of the L2 mental lexicon (Kroll & Tokowicz, 2005; Obler & Goral, 2007). Among them I would like to introduce three key questions that have been considered by researchers when developing their own models, and that might also provide theoretical bases for a viable conceptual framework in the current study.

- 1) Do bilinguals store language information in a single (shared) memory system or in separate systems for each language?

- 2) How is L1 and/or L2 conceptual (or semantic) information interconnected in the bilingual mental lexicon?
- 3) How do bilinguals activate and select relevant lexical items from each language?

These three issues have always been at the centre of the modelling of the L2 mental lexicon.

The sub-sections that follow will cover the attempts to answer the questions, by reviewing early or contemporary research of the bilingual mental lexicon.

3.2.1 Single (Shared) vs. Separate Mental Lexicon

Probably the most significant and enduring issue in the studies of the bilingual mental lexicon is that of whether the bilingual memory of words is stored in a single (shared) system or in separate systems for each language (Kroll, 1993; Kroll & Tokowicz, 2005). This question dates back to 1953, when Weinreich presented an early model characterised by three types of bilingual lexical representations based in Saussurian linguistics: coordinate, compound and subordinative (De Groot, 1993). As shown in Figure 3.1, Weinreich assumed that in the ‘coordinate’ system bilinguals kept two languages and their meanings in separate mental lexicons (e.g. the English *book* and Russian *kníga*; the example from Weinreich [1953/1968]), which means that a word in one language (L1) and its translation in the other language (L2) are represented in two conceptual forms (see Figure 3.1 (a)). On the other hand, in the ‘compound’ system bilinguals had a single underlying conceptual system for L1 and this seemed to be shared by L2, as in Figure 3.1 (b). The third type of bilingual representation (subordinative) can be described as the second language learners’ system being in transition to the compound system (see Figure 3.1 (c)).

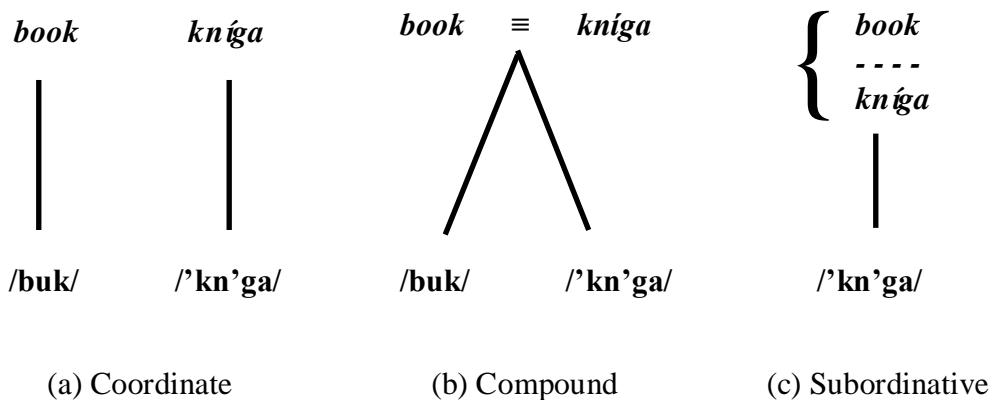


Figure 3.1 Three types of word knowledge in bilinguals (from Weinreich, 1953/1968)

According to Weinreich (1953/1968), it seemed that these three types of organization could co-exist in one individual's bilingual mental lexicon, where their usage might be determined depending on the degree of the linkage of words between L1 and L2. However, this account of three possible systems of bilinguals' mental lexicon seems to lack detailed explanation about how they worked in specific situations and thus to remain limited to verbal descriptions (Thomas & Van Heuven, 2005). For example, it appears that in distinguishing between compound and coordinate systems Weinreich did not consider the proficiency of bilinguals' two languages (De Groot, 1993). Another problem of this early model is that it failed to take into consideration different domains such as orthography, phonology, semantics and syntax (Kroll & Tokowicz, 2005).

However, Weinreich's three types of word knowledge in the bilingual mental lexicon (particularly the first two: coordinate and compound) inspired researchers to attempt to develop various models of bilingual memory storage. These carried similar but different

labels, such as shared storage vs. separate storage hypothesis; interdependence vs. independence hypothesis; single-code vs. dual-code hypothesis (De Groot, 1993). Although the empirical evidence in the literature on bilingual language processing yields conflicting results that support either a shared or a separate model, it seems clear that the results in any specific research will be dependent upon the attributes or components of the bilingual mental lexicon emphasised in that study. In general, previous research supported a single (shared) memory model for the semantic or conceptual domain, whereas it seemed to agree on a separate memory model for other domains of the bilingual mental lexicon (Durgunoglu & Roediger, 1987). As Kroll & Tokowicz (2005: 531) also mentioned, “There is no reason why the answer to the question of how the two languages are represented needs to be the same for orthography, phonology, semantics, and syntax.”

Accordingly, the assumption that bilinguals stored semantic or conceptual information in a shared system led to the emergence of hierarchical models of bilingual memory representation (see Section 3.2.2), whereas the idea that the other lexical information could be stored separately in independent memory systems naturally gave rise to questions related to the language selective and language non-selective hypotheses (see Section 3.2.3).

3.2.2 Revised Hierarchical Model (RHM)

The second theoretical issue in modelling the bilingual mental lexicon is that of how L1 and/or L2 conceptual (or semantic) information is interconnected. Hierarchical models, in general, seemed to agree that bilinguals stored that information in a single (shared) memory system rather than in separate systems for each language. Kroll & Sholl (1992: 192) referred to this situation as “cross-language semantic priming”, because “the two languages access a

common conceptual representation". However, with regard to lexical information from other domains (or levels in the hierarchy) in the bilingual mental lexicon, researchers assumed that different types of mental activities could occur.

We may then ask how the lexical information of L2 words links to concepts in a pre-existing L1 mental lexicon. Potter, So, Von Eckhardt & Feldman (1984) proposed two hypotheses of cross-language connection: word association and concept mediation, as presented in Figure 3.2. The major difference between these two hypotheses is how bilinguals make contact with concepts. According to the word association hypothesis (see Figure 3.2 (a)), bilinguals should always need L1 (translation equivalent) for L2 to have access to conceptual information. On the other hand, in the concept mediation hypothesis (see Figure 3.2 (b)), both L1 and L2 can have direct access to a common conceptual representation. Potter et al. (1984) conducted picture naming and word translation tasks in order to show that both hypotheses were relevant to some extent.

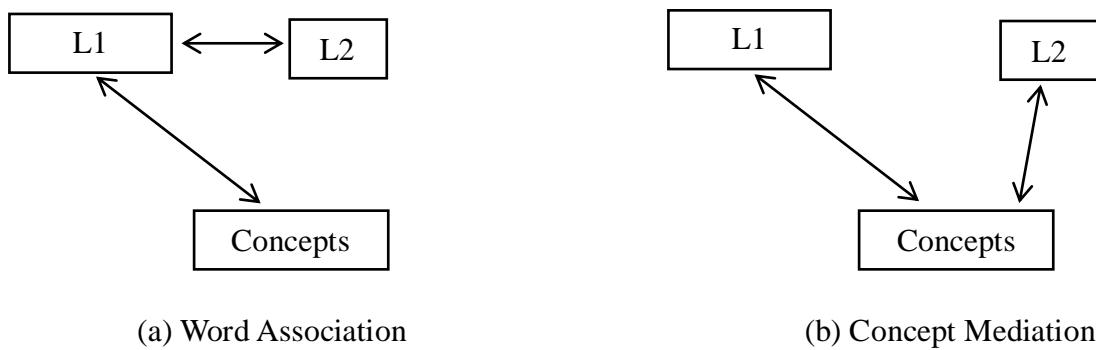


Figure 3.2 Two hypotheses of cross-language connection (adapted from Potter et al. (1984))

Later studies (e.g. Chen & Leung, 1989; Kroll & Curley, 1988) further investigated these two hypotheses, suggesting that L2 learners at a low proficiency level are more likely to follow the word association process known as ‘lexical mediation’, whereas learners at higher proficiency levels may be able to make direct contact with the meaning of L2 words without the help of translation equivalents (L1 information), following the process called ‘conceptual mediation’ (Barcroft & Sunderman, 2008).

Furthermore, Kroll & Stewart (1994) observed the ‘asymmetric directionality’ in cross-language connection, a phenomenon not taken into account in Potter et al.’s two hypotheses, which suggested that the strength of connections between lexical and conceptual mediation could differ according to the direction of translation.

In order to address the developmental shift from lexical to conceptual mediation as well as the asymmetric directionality in translation, Kroll & Stewart (1994) proposed the Revised Hierarchical Model (RHM), which integrated the lexical and conceptual mediation of Potter et al., as presented in Figure 3.3.

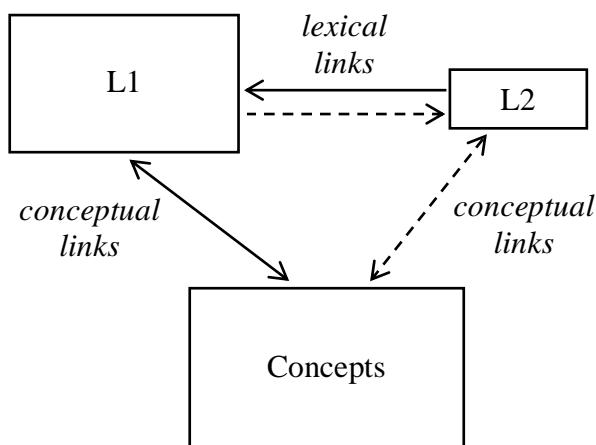


Figure 3.3 Revised Hierarchical Model (from Kroll & Stewart (1994))

The RHM assumed that both lexical and conceptual links are available in bilingual memory and that the links are bidirectional, although their strengths differ according to bilinguals' language proficiency: the solid line denotes strong links between components whereas the dotted line indicates weak links (Kroll & Stewart, 1994). In the RHM, as Kroll & Tokowicz (2005: 546) described: “[T]he learner exploits the existing word-to-concept connections in L1 to access meaning for new words in L2. Thus, a strong lexical connection from L2 to L1 will be established during learning. As learners become more proficient in L2, they will begin to develop the ability to conceptually process L2 words directly.” In other words, they argued that lexical links from L2 to L1 are stronger than those from L1 to L2, and this seemed to be why translation from L1 to L2 takes longer than translation from L2 to L1. They also argued that conceptual links for L1 are stronger than those for L2, suggesting that more proficient bilinguals may be able to access the meanings of L2 words without the help of their L1.

According to Ellis (2008: 375), the RHM has been a very important psycholinguistic model “in emphasizing the interaction in between lexical and conceptual representations in the bilingual lexicon”. Therefore, this model provides important implications for the conceptual framework for the current study.

3.2.3 Language Selective vs. Language Non-selective Hypotheses

The third theoretical issue regarding the bilingual mental lexicon is how bilinguals activate and select lexical items they intend to use. Although it is generally acknowledged that bilinguals share conceptual (or semantic) information in the L2 mental lexicon as discussed in Section 3.2.2, it seems that lexical information from other domains, such as phonology,

orthography, morphology and syntax, is stored separately for each language. This raises the question of how these other domains from each language are connected to the concepts. For example, with regard to comprehension, how do bilinguals access relevant concepts with given lexical items (or stimuli) in order to understand them? In the case of production, how do bilinguals draw upon their intended conceptual information stored in the shared memory system to speak or write relevant lexical items in the target language?

This question has brought forth two hypotheses, one arguing the case for language selective access, and the other for non-selective access (Grainger, Midgley & Holcomb, 2010). De Bot (2004: 18) summarised these two with the question: “when we are confronted with a word, [...] do we first access the lexicon from one language and then the next [language selective], or is there a parallel search through all languages, words not being organised primarily through language [language non-selective?]”

The language selective hypothesis is based on the idea that bilinguals use a switching mechanism that selects a target language for language processing. It assumes that more proficient bilinguals are able to ‘switch on’ a relevant language in a given context, often rapidly and intentionally, while they can simultaneously ‘switch off’ the other language (Meuter, 2005). If this is the case, it seems that bilinguals send information extracted from the stimulus directly to the target lexical storage in order to find the appropriate set of language-specific representations (Grainger et al., 2010).

In contrast, the language non-selective hypothesis assumes that bilinguals select the most appropriate representation among many candidates called ‘language nodes’, which are activated by the stimuli from both languages (De Bot, 2004) (for more information about language nodes, see Section 3.3.2). This decision-making process is called ‘lexical selection’.

According to Costa (2005), the language nodes are competing against each other to be selected. He suggested that language nodes that are selected by bilinguals include lexical information from domains such as phonology, orthography and syntax in the bilingual mental lexicon.

Most of the early studies demonstrated so-called ‘switching cost’ as evidence for the language selective hypothesis (e.g. Grainger & Beauvillain, 1987; Thomas & Allport, 2000). The switching cost can be described as a price that is paid when bilinguals alter languages in the L2 mental lexicon (Meuter, 2005). For example, Grainger & Beauvillain (1987) administered lexical decision tasks with English-French bilinguals, and showed that the reaction time was faster with pure-language lists than with mixed-language lists. This may suggest that bilinguals had to pay a switching cost by spending more time when they dealt with mixed-language lists.

However, research that supports the language non-selective hypothesis argues that the switching cost can also be interpreted as ‘inhibitory effects’ of the non-target language when both languages are being activated at the same time (De Groot, Delmaar & Lupker, 2000; Grainger & Dijkstra, 1992). As Grainger et al. described:

In comprehension, switch costs arise from bottom-up activation of a given language node driven by presentation of a word in that language, leading to inhibition of lexical representations in the other language. In production, the appropriate language node is activated top-down in order to ensure that only lexical representations in the target language are selected for output.

Grainger et al., 2010: 271

Cross-language interference between two languages offers more direct evidence for the language non-selective view, because it strongly suggests that both languages are activated, with a non-target language inhibiting the target language. On the other hand, it appears that the language selective hypothesis cannot offer an appropriate explanation for the interference in the bilingual mental lexicon (Grainger et al., 2010). However, those who argue for language non-selective access still need to explain how some bilinguals are not hindered by a non-target language when they use a target language (cross-language interference).

Therefore, it appears that the debate between the language selective and language non-selective hypotheses is still ongoing (Obler & Goral, 2007). However, rather than focusing upon this dichotomy, researchers have raised other more complex questions, such as: ‘Does language proficiency affect the selection process?’ ‘How is the selected language maintained?’ (Meuter, 2005)

3.2.4 Perspectives of the Current Study toward the L2 Mental Lexicon

Based on the above reviews of the three theoretical issues considered by previous research with regard to the bilingual mental lexicon, I would like to present the perspectives adopted by the current study toward those issues.

Firstly, with regard to the question of single or separate storage, this study basically follows the RHM in assuming that bilinguals have a shared memory system for conceptual (or semantic) information while, for all other lexical domains, they use separate memory systems for each language. As Barcroft & Sunderman (2008: 279) indicated: “It [RHM] proposed independent lexical representations for words in each language, but an integrated conceptual system.”

Here it is necessary to clarify what is meant by the ‘conceptual’ or ‘semantic’ information referred to in previous research. While the two terms may seem to be synonymous, the meanings of both are quite equivocal, and consequently many researchers have not always been careful to distinguish one from the other (Jarvis, 2009). However, for the purposes of this study it will be helpful to draw a line between them, in order to categorise L2 learners’ errors by lexical domains. According to Jarvis (2009: 100): “Concepts reflect the level of thought and experiential knowledge, and they consist of various types of mental images, image schemas, mental scripts and forms of knowledge that are organized into structured categories of thought and categories of meaning.” On the other hand, semantic information can be described as ‘lexicalised concepts’ that are linked to conceptual information. In other words, concepts may exist in the forms of ‘images’, but should be verbalised (or lexicalised) with particular lexical items. Therefore, in the current study, conceptual information is regarded as superordinate, mapping to language-specific semantic information (L1 or L2) that belongs to one of the lexical domains (see Jiang’s L2 vocabulary acquisition model in Section 3.3.3). I consider this distinction to be relevant because this study deals with lexical errors produced by adult L2 learners. As Kroll (1993: 55) also argued: “For adults who already have a fluent and dominant first language, and for whom the second language is acquired within the cultural context of the first language, the problem is not to learn new concepts, but rather to acquire new mappings between concepts and the second language words.”

Therefore, the current study assumes that a particular lexical item is made up of different pieces (or hierarchies) of knowledge from lexical domains such as semantics, syntax, morphology, orthography and phonology (Kroll & Tokowicz, 2005). Since bilinguals use a shared (single) memory system for concepts which have probably the strongest connections with semantic information across both languages, the semantic domain in the L2 mental

lexicon probably plays the most important role as a bridge over other domains between languages (De Bot, 2004; De Bot & Schreuder, 1993). Furthermore, this suggests that if bilinguals (or L2 learners) produce some errors in L2 production, the errors can perhaps be categorised by the lexical domains that make up the L2 mental lexicon, one of the two criteria for classifying L2 learners' errors in this study (see Section 3.4.2).

Secondly, the current study adopts the perspective of language non-selective access, where bilinguals simultaneously activate language nodes from both languages and select the most appropriate ones in the lexical domains through a competing process with plausible candidates. The decision to adopt this hypothesis has been made not only because the language non-selective view seems to be more favoured in recent research (Grainger et al., 2010), but also because this perspective supports the possibility of interference between the target and non-target languages in the course of 'lexical selection' (Costa, 2005). Thus, it entails the second error criterion in the new taxonomy proposed by this study, namely the error sources (L1 or L2) (see Section 3.4.2). Note that the previous research has used a number of terms to describe the mental processes in 'lexical selection', for example lexical processing, lexical decision, and decision-making process. The current study will use the term 'cognitive vehicles', which encompasses those mental or cognitive processes (see Section 3.4.2).

Finally, it should be noted that language processing procedures might differ according to whether the L2 mental lexicon is involved with language comprehension or with language production. As already briefly mentioned in Section 3.2.3, comprehension is a primarily bottom-up process whereas production is top-down (Costa, 2005; Grainger et al., 2010). Channell (2013) described these two processes as 'sound to meaning direction' and 'meaning to sound direction' respectively, because in comprehension the sound of words is decoded to

search for the corresponding meaning in the L2 mental lexicon, whereas in production the process starts from the meaning in order to select appropriate words that are then converted into phonological forms. Nation (2001) also discussed the receptive/productive distinction in terms of ‘knowing a word’, which involves various kinds of language knowledge and use. He presented a model that focused on parts of language knowledge and its use, indicating that ‘knowing a word’ means knowing three parts of the word (i.e. form, meaning and use), each of which can be subcategorised by two kinds of vocabulary knowledge: receptive (comprehensive) and productive knowledge. He also suggested that L2 learners would usually face a greater learning burden in production than in comprehension. Although he focused more on language knowledge and use in his model, he admitted: “It is also possible to show the aspects of what is involved in knowing a word using a ‘process model’, which emphasises the relations between the parts” (Nation, 2001: 26 [single quotation marks added]).

Accordingly, when researchers have developed their own language processing models for the bilingual mental lexicon, those models have usually been specific to either language comprehension or language production. Regrettably, far less research has examined production with regard to the bilingual mental lexicon than has investigated comprehension (Kroll & Tokowicz, 2005).

Therefore, the next section will review language production models in previous research, which might be significant as theoretical bases in order to present a conceptual framework for a new error taxonomy.

3.3 Language Production Model

In order to discover the mental processes involved when language users make lexical errors, it is necessary to account for the components of the mental lexicon and how these work. If one can explain each step of the language production process in the mental lexicon, it may then be possible to describe how the lexical errors are produced and which elements in the mental lexicon cause the errors. This section introduces three language production models proposed in previous studies. It aims to arrive at a language production model that is relevant for this research and to specify the causes of lexical errors, which can be detected in some language processing stages.

A number of cognitive linguists have speculated as to the language production process. Among the many language production models that have been presented, three are particularly relevant to the current study:

- 1) Levelt's (1989, 1999) L1 speech production model
- 2) De Bot's (1992, 2004) multilingual production model
- 3) Jiang's (2000, 2004) L2 vocabulary acquisition model

These three early models each provide a theoretical basis for understanding mental processes in language production. In this study they are used to inform the development of a language production model that will form a conceptual framework to describe lexical errors produced by L2 learners.

3.3.1 Levelt's L1 Speech Production Model

Levelt's (1989) L1 speech production model is a very good starting point to discuss language production procedures, because at the time of its introduction it was "the most complete, empirically based language production model available" (De Bot, Paribakht & Wesche, 1997: 310). Levelt proposed the L1 speech production model as "a blueprint for the speaker" (Levelt, 1989: 9). He assumed that a human has an innate articulatory system, which he characterised as "a highly complex information processor" consisting of five language processing components as shown in Figure 3.4: conceptualiser, formulator, articulator, audition and speech comprehension system. Among the five components, the first three are mainly involved in language production.

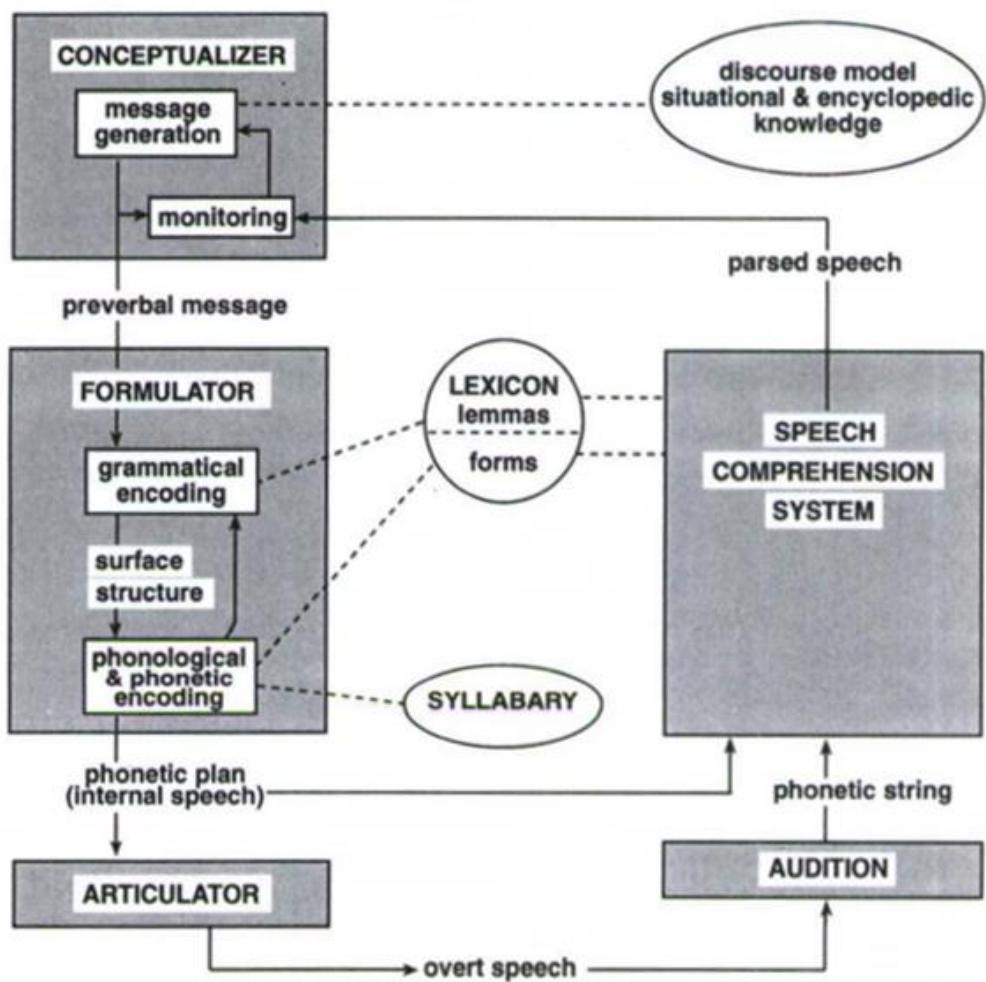


Figure 3.4 A blueprint for the speaker (Levelt, 1989: 9)

Firstly, the conceptualiser generates messages that are influenced by a speaker's declarative and procedural knowledge (Levelt, 1989). The output of the conceptualiser is a preverbal message, which in turn becomes an input of the second component, the formulator. Secondly, the formulator retrieves sources of lexical items stored in the mental lexicon, in order to encode semantic, syntactic, morphological and phonological information. One important theoretical assumption about how these lexical items are stored in the mental lexicon is that the properties of a lexical entry are embedded in two different areas: lemma and lexeme. The

terms lemma and lexeme are commonly used in the field of lexicography (Jarvis, 2009). The lemma contains semantic and syntactic properties of a lexical item, whereas the lexeme (also known as ‘form’) contains morphological and phonological information. Although Figure 3.4 does not clearly show how semantic information of a lexical item is treated, it is generally assumed that the semantic encoding is also processed in the formulator, because a lexical item is retrieved using its meaning (De Bot, 1992). In other words, when the conceptualiser activates conceptual information, the formulator links the concepts to semantic features in the lemma. The output of the formulator is called the phonetic plan (internal speech), which is not yet overt speech. Lastly, the phonetic plan is executed by the articulator in the form of overt speech.

Levelt’s speech production model was very influential. Several researchers have adopted the model to investigate various aspects of monolingual as well as bilingual performance (e.g. De Bot, 2004; Green, 1993; Poulisse, 1999; De Bot & Schreuder, 1993). It is significant mainly in its conceptualisation of a language production device from a cognitive linguistic perspective and in its description of how language is produced within the components of that device. Although the model focuses on L1 speech processing, it could also provide insightful ideas about L2 acquisition and the production of written forms of L2. As De Bot (1992: 2) argued, “because every unilingual speaker has the potential to become bilingual, the validity of a model can be tested by examining whether it is suitable for bilingualism”.

3.3.2 De Bot’s Multilingual Production Model

De Bot (2004) altered Levelt’s model to account for bilingual production and further for multilingual production. He then proposed ‘the multilingual processing model’ as shown in

Figure 3.5. He aimed to keep the original Levelt model, while changing certain parts based on empirical findings of bilingual research. In order to alter the L1 production model to a bilingual or multilingual model, a few assumptions were necessary. First of all, De Bot (1992) assumed that there is a single language storage system for conceptual information, because the conceptual information itself is not language specific. Furthermore, with regard to the formulator and mental lexicon of Levelt's model, it could be that there is a separate formulator and a separate lexicon in each language, or one large storage system that contains all the multilingual information about two (or more) languages. De Bot assumed that the semantic (meaning) part of the lemma could be shared by more than one language, whereas the syntactic part and the lexeme (morpho-phonological information) could be separate for each language. Based on neuro-linguistic research findings, he adopted the 'Subset Hypothesis', which assumes a 'single' mental lexicon where elements from each language can be strongly or weakly linked. In other words, some elements in one language are strongly linked to those in another language (and some are not), which results in the formation of subsets in the mental lexicon (De Bot, 1992). Accordingly, De Bot's model seems to be in line with the same perspective that the current study adopts with regard to the storage of L2 mental lexicon, as discussed in Section 3.2.4.

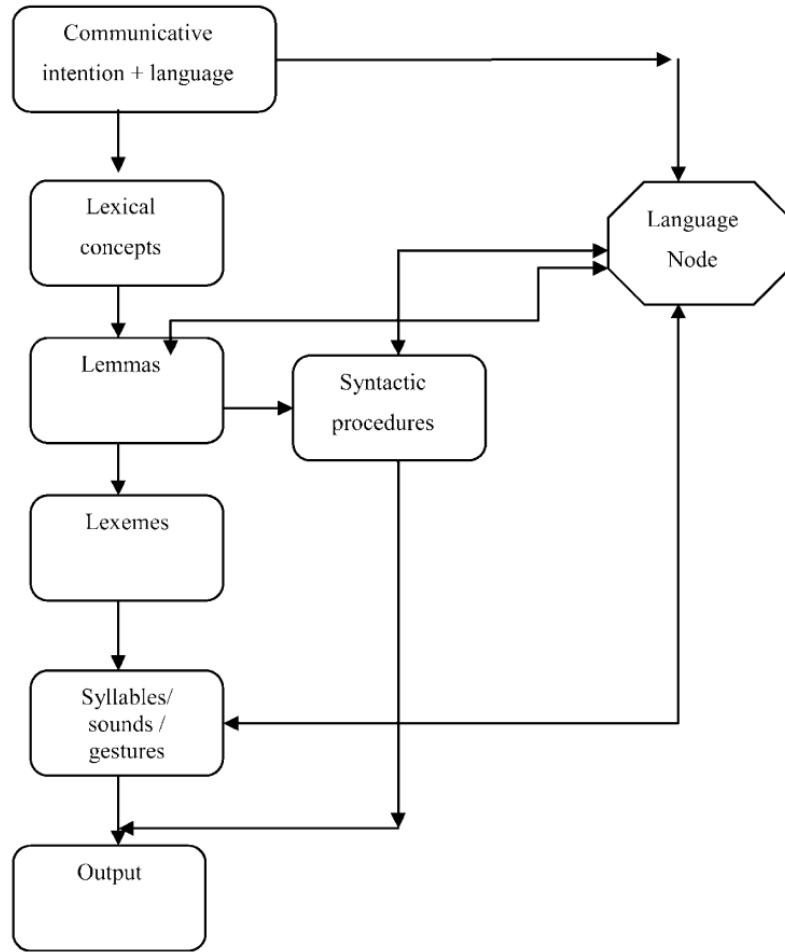


Figure 3.5 The multilingual processing model (De Bot, 2004: 29)

As can be seen from Figure 3.5, the multilingual processing model has a mediating component, the ‘language node’ (see Section 3.2.3). De Bot (2004: 28) argued: “The language node controls the various processing components with respect to the language to be used [...] When a particular language is called for, the language node will inform all relevant components, that is those components in which syntactic or form information needs to be selected, about the subset to be activated.” Therefore, in De Bot’s multilingual model, the lexical information to be produced comes from the language nodes, which are activated by

lemma and lexeme from multiple languages, whereas in Levelt's L1 production model there is no need for the language node.

De Bot's multilingual processing model is noteworthy in that it is not confined to L1 production but attempts to show the L2 production procedures by presenting their relevant components. Since De Bot (2004) claimed that this model was developed based on a wealth of empirical evidence provided by a number of studies, such as cross-linguistic, eye-tracking, brain-imaging and language production studies, it is possible that at the time of its introduction the model offered a more accurate explanation of L2 production procedures than had previously been achieved.

3.3.3 Jiang's L2 Vocabulary Acquisition Model

Jiang's (2000) L2 vocabulary acquisition model is notable because it provides what constitutes the L2 mental lexicon. The model presents "an account of psycholinguistic processes involved in lexical presentation and development in L2" (Jiang, 2000: 72), and this will help explain how and why lexical errors are made by L2 learners and is closely related to the error taxonomy that will be presented in Section 3.5.2. It focuses particularly on L2 learners' mental lexicon and L2 lexical development and representation and assumes that the L2 lexical representation is fundamentally different from that of L1.

Based on Levelt (1989), Jiang (2000) assumed that the mental lexicon can be divided into two components: lemma and lexeme, where the lemma contains semantic and syntactic information, while the lexeme contains morphological and formal (phonological and orthographic) information about a lexical item (see Figure 3.6).

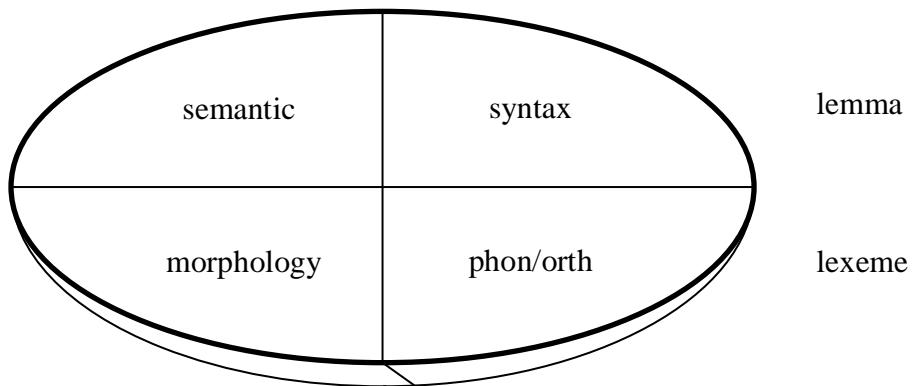


Figure 3.6 The lexical representation of the mental lexicon (Jiang, 2000: 48)

Arguing from a cognitive linguistic perspective, he asserted that there are three stages of L2 vocabulary acquisition, on the basis of what is represented by the lexical entry, as shown in Figure 3.7.

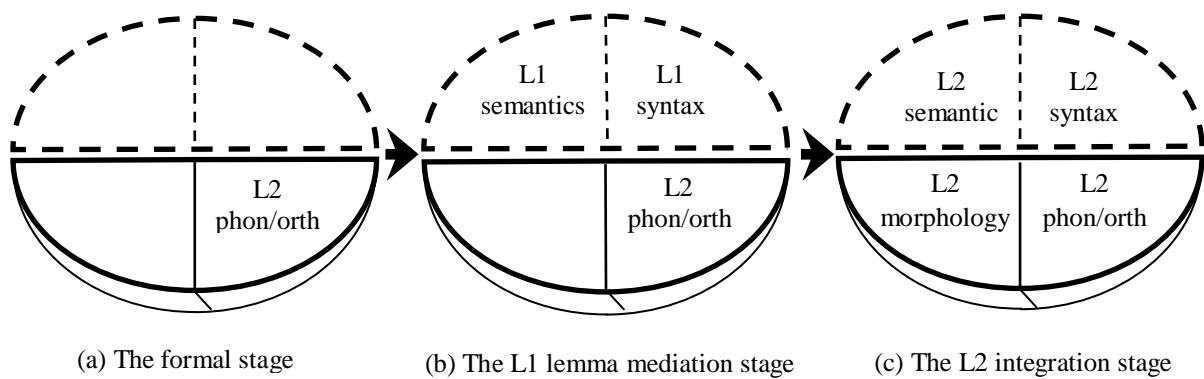


Figure 3.7 The three stages of L2 vocabulary acquisition (Jiang, 2000: 54)

In the first, ‘formal’, stage, lexical entries are established phonologically and orthographically. Although these lexical entries do contain some semantic, syntactic and morphological

information (shown as dotted lines in Figure 3.7 (a)), this may work only as a ‘pointer’ linking L2 words to the corresponding L1 translations.

In the second, ‘L1 lemma mediation’, stage, L2 learners borrow the lemma information (semantic and syntactic specifications) from the L1 counterpart, mediating L2 word use (see Figure 3.7 (b)). Since L2 learners already have their own L1 lexical system, they are likely to rely on this in the L2 acquisition process, rather than creating a separate L2 lexical system. Hence, it seems that in this stage L2 learners resort to L1 translations for meaning. However, the semantic specifications between L1 and L2 do not perfectly overlap and this might be why various kinds of semantic errors occur. One possible difference between the characteristics of the first and the second stages is the strength of the connection between L2 words and L1 translations. If the connection is strong, then L2 learners will be willing to use L2 words to a greater extent, although the semantic specification (meaning) could be slightly different from the original meaning. For example, the word *house* can be translated by the Korean word *cip* 之家, so that Korean learners of English might carry the semantic concept of Korean *cip* over to the English word *house*. However, the shape of the Korean *cip*, which Koreans can easily picture, is quite different from that of the English *house*. In order for Korean learners of English to fully acquire the original meaning of *house* they might have to go through the process of ‘semantic restructuring’ (Jiang, 2004). This kind of disparity, which I will call ‘incongruity’, is associated not only with semantic representations, but also with syntactic, morphological and even orthographic representations, although Jiang did not seem to be sure whether morphological and orthographic specifications are affected by L1 or L2. This concept of incongruity between L1 and L2 informs the new error taxonomy proposed in Section 3.5.2.

Finally, in the third, ‘L2 integration’, stage, the L1 lexical information is integrated into the L2 lexical entry (see Figure 3.7 (c)). Learners who reach this stage can use English with native-like proficiency. However, Jiang argued that the majority of L2 learners remain in the second stage.

Almost all lexical errors are made in the first and the second stages, where there are mismatches in terms of semantic representations, for example a lack of semantic overlap between L1 and L2. As Jiang (2000: 57) pointed out, “words in two different languages seldom share identical semantic specifications. Their syntactic properties may differ as well. Under both circumstances, lexical association and L1 lemma mediation will often lead to lexical errors.”

In this context, Jiang (2000: 67) divided L2 words in the mental lexicon into three categories according to the degree of semantic overlap: *real friends*, *false friends* and *strangers*. The L2 words that belong to the *real friends* category have a high degree of semantic overlap, while *false friends* are words that have a low degree of semantic overlap. The *strangers* category contains the L2 words for which learners have only the formal (phonological and orthographic) information in the mental lexicon but not the corresponding L1 translations (copied lemma). L2 learners would probably not make many errors with the *real friends*, where the L2 semantic information is almost identical to that of L1 equivalents. However, they may produce more errors with the *false friends*, presumably because of L1 transfer (interlingual influence), where the semantic information that is copied from L1 could interfere with the L2 lexical development. On the other hand, L2 learners might avoid using the words classified as *strangers*. If they do attempt to use these words, their lack of semantic information might result in errors that are peculiar and difficult to predict.

Regrettably, in the division of L2 words Jiang described only the degree of ‘semantic overlap’ between L1 and L2. However, I would argue that there are similar kinds of incongruity and overlap between L1 and L2 in the ‘syntactic’, ‘morphological’ and even ‘orthographic’ domains. Therefore, the current study focuses on how L2 learners behave when they are confronted with these kinds of incongruity (i.e. the given situations where they lack lexical and grammatical knowledge of L2 words) and proposes a new L2 lexical development model to explain specific types of lexical errors from each domain in the L2 mental lexicon (see Section 3.5.1).

3.3.4 Contributions of the Three Language Production Models to the Current Study

Previous three sub-sections have reviewed three language production models in order to present a relevant conceptual framework within which this study can describe how L2 learners’ lexical errors are produced. These models provide insights that are useful to develop the L2 lexical development model in the current study (see Section 3.5.1).

First, Levelt’s (1989) L1 speech production model was a good starting point for the discussion of the generation of L1 speech, because it provided the fundamental procedures of language production with the processing components. Secondly, De Bot’s (1992) bilingual (or multilingual) production model expanded the notion of a language processing framework to cover L2 production by adding the mediating component ‘language node’, which is believed to control the other processing components between L1 and L2. Finally, Jiang (2000) showed that L2 lexical representation is significantly different from L1 representation. Jiang’s L2 vocabulary acquisition model described the process of L2 vocabulary acquisition in three stages.

Jiang's L2 vocabulary acquisition model, in particular, makes two very important contributions to the current study. First, it provides a theoretical basis for L2 learners' mental lexicon. According to Jiang, the lexicon is affected by L1 and L2 and comprises four lexical domains: semantic, syntactic, morphological and phonological/orthographic. These domains are one of the two criteria used in the current study to create a new error taxonomy (see Section 3.5.2), where the lexical representations from the four domains result in four different kinds of lexical components of L2 learners' errors (see Section 3.5.2). Second, Jiang's model, like the other two models discussed above, highlights the role of L2 learners' mental processes in producing lexical representations. In previous EA studies, researchers also noted these mental processes and sometimes used them as important criteria to classify errors within certain categories, e.g. communication strategy-based errors (see Section 2.1.2). However, the previous EA studies seemed to be inconsistent in their treatment of mental processes and this often resulted in ambiguous error taxonomies. Therefore, in the current study, I will use the term 'cognitive vehicles' to encompass the mental processes that are active in the course of L2 acquisition (see Section 3.5.1). As will be pointed out in Section 3.5.1, it is important to note that the function of the cognitive vehicles should be excluded from the error criteria in the proposed L2 lexical development model.

3.4 Sources of Lexical Errors

3.4.1 Diagnostic Approach to Lexical Errors

As mentioned in Section 2.1.4, the possible causes of lexical errors are an appropriate criterion to establish a new error taxonomy. Describing errors based on their surface structure is different from diagnosing the causes of errors. As Dulay et al. (1982) stated, "the accurate

description of errors is a separate activity from the task of inferring the sources of the errors”.

The reason for attempting to discover the causes of learners’ errors is that this is expected to provide more pedagogic implications than a (surface) structure-based taxonomy. As mentioned in Sections 2.1.4 and 2.1.5, some previous studies have used possible causes as criteria to divide lexical errors according to their own classifications. However, the error taxonomies developed through this diagnostic (cause-based) approach seem to be based on somewhat ambiguous error criteria, probably because they depend upon the subjective viewpoints of individual researchers, as mentioned as one of the limitations of EA (see Section 2.1.1).

For example, Richards (1974) used a diagnostic approach to divide lexical errors into three categories: interference, intralingual and developmental error. Interference error (also known as L1 transfer error) is caused by learners’ L1, while intralingual errors arise within the structure of the target language. Developmental error occurs when learners attempt to develop their own hypotheses about the target language from their limited experience. However, the domain of developmental error is not clear-cut but seems to overlap with those of interference and intralingual error. According to Schachter & Celce-Murcia (1977: 443): “The [developmental] errors show that the learner – oftentimes completely independent of his native language – is making false hypotheses about the target language based on limited exposure to it.” This suggests that developmental error is not always dependent on L1. Indeed, when the learner makes false hypotheses about the target language this involves L2, sometimes with no input at all from L1.

Similarly, James (1998) introduced four main diagnosis-based categories of lexical errors: interlingual, intralingual, communication strategy-based and induced error. The interlingual and intralingual errors are identical to Richards’ (1974) interference and intralingual error

respectively. On the other hand, the communication strategy-based error is produced when L2 learners resort to communication strategies, such as circumlocution or approximation. Communication strategy can be defined as “a systematic attempt by the learner to express or decode meaning in the target language, in situations where the appropriate systematic target language rules have not been formed” (Tarone, Cohen & Dumas, 1983: 5). Meanwhile, induced error “results more from the classroom situation than from either the students’ incomplete competence in English grammar (intralingual errors) or first language interference (interlingual errors)” (Stenson, 1983; as cited in James, 1998: 189). However, again, James’s diagnosis-based categories are not clear-cut. For example, the subcategories of communication strategy-based error include language switch (borrowing) and calque (literal translation), which are prototypical examples of interlingual error. Borrowing is the phenomenon whereby L2 learners directly use (or ‘borrow’) L1 words, rather than translate them. Calque occurs when L2 learners literally translate words or phrases from the L1. According to scholars such as Odlin (1989) and Ringbom (2001), L2 learners from different L1 backgrounds made calque errors. Another subcategory of communication strategy-based error, circumlocution, could also be categorised as intralingual error. In addition, induced errors might be a kind of intralingual error because they are caused by L2 input such as teaching materials or teacher talk.

Many other studies have applied ambiguous and overlapping criteria when categorising L2 learners’ lexical errors according to possible causes. Carrió & Mestre (2014) studied lexical errors in scientific papers written in English by L1 Spanish researchers. They divided the lexical errors into three main categories, each containing three to five subcategories based on the causes of the errors. In addition to interlingual and intralingual errors, traditionally considered as the main error types, Carrió & Mestre (2014) added a third category,

‘conceptual error’. Based on their empirical data analysis, the authors claimed that ‘conceptual error’ is “caused by the failure of the speaker to match an idea with the correct expression, i.e. a breakdown of the concept-term relationship” (Carrió & Mestre, 2014: 99). Unfortunately this categorisation shares the same major weakness as seen in previous studies, namely that the three main categories appear to overlap. For example, one of the subcategories of conceptual error is ‘use of a word due to confusion over meaning’. The meaning (concept) of a word in this subcategory is based not only on the learners’ L1 but also on the L2 in the mental lexicon, which implies that this kind of error could also belong to either the interlingual or intralingual error category. Similarly, while ‘erroneous collocation’ appears as a subcategory under intralingual error, collocational errors could also be classified as interlingual errors, according to the possible causes (see Nesselhauf, 2005).

The unclear and overlapping criteria in the diagnosis-based error taxonomies of previous studies give rise to certain reasonable questions: What causes this ambiguity in error categories? How might one make the categorisation clear? Based on the reviews of theoretical issues around the L2 mental lexicon and language production models in Sections 3.2 and 3.3, the current study argues that the ambiguity arises out of inconsistent application of equivocal error criteria. In order to solve these problems, the next section will attempt to explain some of the terms that I consider to have been used without clear definition in the previous studies. Then, Section 3.5.2 will propose a new taxonomy based on a refined conceptual framework of L2 acquisition.

3.4.2 Sources of Errors and the Role of Cognitive Vehicles

The reviews of previous studies on diagnosis-based error taxonomies indicate certain areas of agreement with regard to possible causes of errors. First, the previous studies all pointed to two major groups of factors involved in causing lexical errors, namely interlingual and intralingual factors. Second, most of the studies also presented a third (or fourth) error factor, which indicates that researchers generally agreed that there are other factors involved, albeit that they differed in their views as to what these may be. These third (or fourth) error types can be grouped into two kinds, namely input-related (L1 or L2) error, such as induced error; and mental process-related error, such as communication strategy-based error. However, from a cognitive linguistic perspective to L2 mental lexicon, it seems quite clear that the ‘sources’ of errors can be narrowed down only to two: L1 and L2 information in bilingual mental lexicon. On the other hand, the other factors that previous studies claimed may belong to these two sources or seem to be like cognitive processes, not the ‘sources’ themselves. In other words, the current study sees the language store and cognitive processes as two different things (see Sections 3.2 and 3.3).

I would argue that the lack of clarity in categorisation in the previous diagnosis-based error taxonomies was a result of the rather vague use of the terms ‘source’, ‘factor’ and ‘cause’. These three terms, used interchangeably (or implicitly assumed as interchangeable) in the previous studies, are seen to have had two different meanings depending on the contexts in which they were used. One meaning is a ‘storing place’ or ‘repository’ that contains the knowledge of word resources; that is, the mental lexicon where knowledge of L1 words co-exists with knowledge of L2 words, with a strong or weak connection between them. For example, literal translation errors are considered to be interlingual errors, because the incorrect words or phrases are literally translated from their equivalents in the L1 store. The

other meaning of the three terms is an ‘affecting factor’ in the course of the mental processing when producing the target language. This meaning is based on the assumption that there is a cognitive processing that affects L2 learners when they are retrieving words in the mental lexicon. As Færch & Kasper (1986: 49), who emphasised the mental process in L2 vocabulary acquisition, suggested, “language transfer is primarily a decision-making procedure rather than an automatic process”. Consequently, these cognitive processes can also be involved in the production of lexical errors. For example, the affecting factor or force of a communication strategy-based error is the force that triggers the error during the cognitive process, rather than the storing place of vocabulary.

In order to clearly differentiate the two separate meanings of the terms, i.e. a ‘storing place’ and an ‘affecting factor or force’, I propose to employ the terms that have been commonly used in cognitive psychology: declarative knowledge and procedural knowledge respectively. According to Anderson (1976), declarative knowledge refers to a language user’s underlying knowledge of facts about the world, whereas procedural knowledge is the knowledge about how to do something in the performance of mental acts. To show the distinction between declarative and procedural knowledge Anderson used an analogy from computing, where declarative knowledge is the data and procedural knowledge is the program used to retrieve that data. That is to say, it appears that the ambiguity of the error categories in previous studies results from the fact that these two types of knowledge (storing place or affecting factor) were used interchangeably in the classification of the error features.

Therefore, in the current study, I will use the term ‘source’ to refer to declarative knowledge of words in the mental lexicon, which is the storing place of vocabulary. For procedural knowledge, I will use ‘cognitive vehicle’ as an umbrella term to cover any kind of mental process or affecting force that has access to the source and brings relevant declarative

knowledge. The cognitive vehicles encompass communication strategies (Tarone et al., 1983), learning strategies (Corder, 1983) and cognitive aspects of language production (Carrió & Mestre, 2014), among others. In addition, I will use the terms ‘factor’ and ‘cause’ interchangeably to refer to anything that affects an event, decision or situation.

As mentioned in Section 3.2 and 3.3, the current study takes models and ideas from a cognitive linguistic perspective to L2 mental lexicon. As opposed to the previous studies that claimed several error sources, the current study has clarified two concepts, i.e. ‘source’ and ‘cognitive vehicle’, which provide a new perspective to analyse lexical errors. In conclusion, the dual approach to the error sources (i.e. interlingual and intralingual influences) is one of the two criteria to establish the new error taxonomy, along with the four lexical domains from the L2 vocabulary acquisition model (Jiang, 2000) (see Section 3.3.3). However, the cognitive vehicles are not considered to be a main source of L2 learners’ errors, because they rather help to trigger errors from the main sources (L1 or L2).

3.5 The L2 Lexical Development Model and the New Error Taxonomy

3.5.1 L2 Lexical Development Model in the Current Study

Based on the L2 vocabulary acquisition model (Jiang, 2000) discussed in Section 3.3.3 and on the diagnostic approach to error classification, I would like to propose the ‘L2 lexical development model’ and suggest a design for a new error taxonomy. Figure 3.8 presents the L2 learners’ mental lexicon within the L2 lexical development model. Like Jiang, the current study assumes that the L2 learners’ mental lexicon consists of four lexical domains, which lend themselves to the four lexical representations: semantic, syntactic, morphological and phonological/orthographic. The reason why the mental lexicon is delineated by dotted lines is

that in the first two stages of development it is not yet fully fledged. If it is developed to the third stage (the L2 integration stage) according to Jiang's model, then the lines should become solid as the learners reach native or near native-like proficiency.

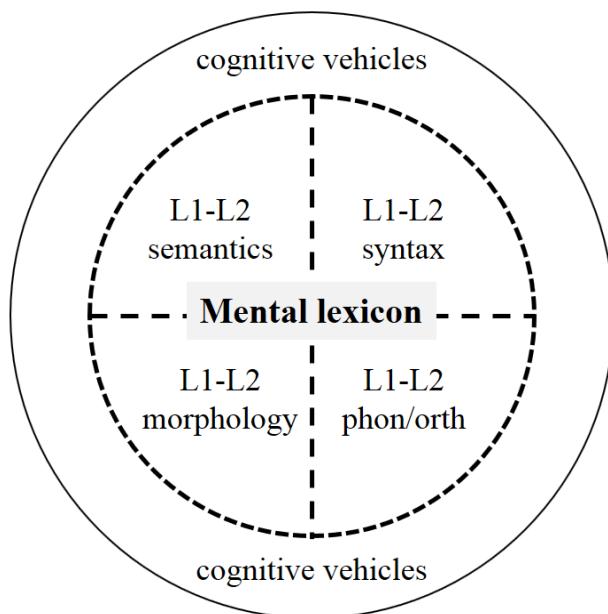


Figure 3.8 The L2 learners' mental lexicon based on the L2 lexical development model

In the current model, unlike Jiang's, I assume that the declarative knowledge in each domain is constituted by a mixture of L1 and L2, where each lexical item can have a different degree of connection between L1 and L2 lexical knowledge. Consequently, the lexical errors could originate not only in L1 but also in L2. It will be very useful to discuss the source-based error categorisation based on the current model, because it is designed to show both error sources (L1 and L2), whereas Jiang (2000) discussed only the interlingual errors caused by learners' L1. Another difference is that Jiang (2000) described the duplication of L2 lemma information (semantic and syntactic knowledge) to the existing L1 conceptual system, but not

the duplication of L2 lexeme information (morphological and phonological/orthographic knowledge). According to his explanation, L2 learners might be able to fully acquire phonological/orthographic knowledge when they first learn the given words in the first stage. In Jiang's view, L1 and L2 morphological information are separated in the mental lexicon and therefore do not interfere with each other, because morphological knowledge is usually considered language-specific (see Section 3.3.3). However, the current model assumes that the L2 lexeme information is also shared by the existing L1 system, just like L2 lemma information. This assumption will be supported by concrete evidence presented in later chapters (see Chapters 9 and 10).

The other major distinction between the L2 lexical development model and Jiang's is the cognitive vehicles that surround the mental lexicon. As mentioned in Section 3.4.2, the cognitive vehicles encompass all kinds of mental processes that activate the sources (L1 or L2) in each lexical domain to generate lexical representations. However, while the cognitive vehicles certainly play an important role in producing lexical errors, they cannot be among the criteria to determine the error categories because they are not a main source of the errors.

Based on the L2 lexical development model in Figure 3.8, in the next section I establish a new error taxonomy. This is based on two main criteria: the four lexical domains of the L2 lexical developmental model and the two error sources (L1 and L2). The lexical errors can be divided by lexical domain into four categories: semantic, syntactic, morphological and phonological/orthographic errors. Then, each category can be divided into two subcategories according to the sources (L1 and L2) that are activated by cognitive vehicles, which finally results in eight subcategories of lexical errors.

3.5.2 A New Categorisation Based on Multi-faceted Criteria

Based on the L2 lexical development model (see Section 3.5.1), Figure 3.9 shows how the lexical errors can be produced. The L2 learners' mental lexicon consists of four lexical domains: semantic, syntactic, morphological and phonological/orthographic.

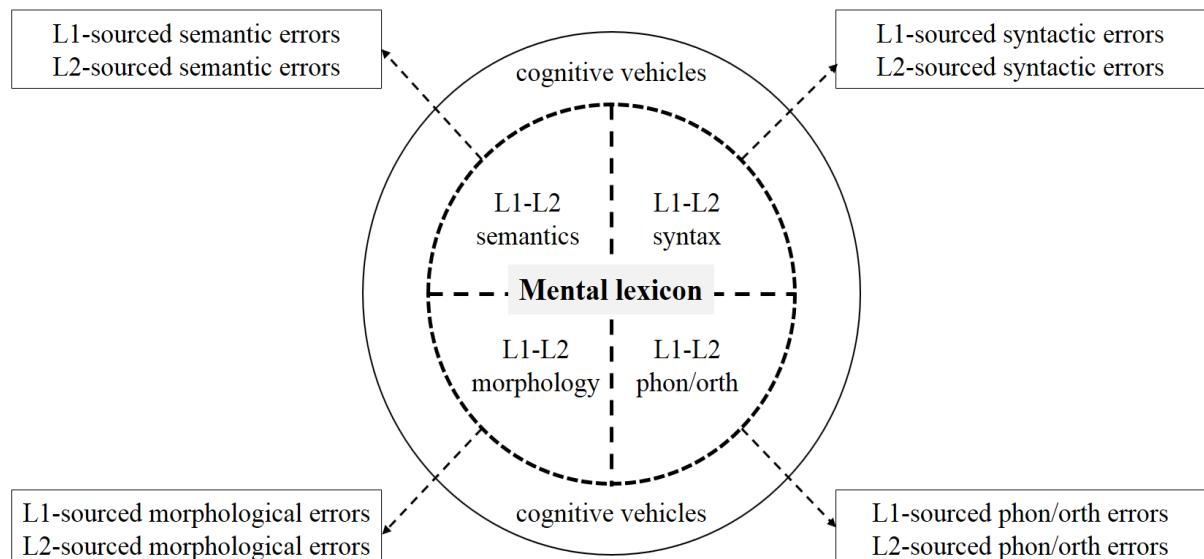


Figure 3.9 The lexical error production

Each lexical domain contains its lexical knowledge in the form of the mixed sources of L1 and L2. Consequently, L2 learners would face ‘incongruity’, where their L1 lexical knowledge of the four domains might not perfectly match with their L2 lexical knowledge.

Table 3.1 presents the new error taxonomy based on the L2 lexical development model.

For example, semantic errors might be affected by an L1 or L2 source. The term ‘semantic errors’ refers to semantically ‘deviant expressions’ of lexis produced by L2 learners. When L2 learners mistakenly choose odd words or strange multi-word units (or collocations), the

resulting expressions are often quite unlike native or conventional usage and might not be found in reference corpora. It is likely that native speakers would consider them as ‘deviant’ and not commonly acceptable English, even though they might be grammatically correct. However, as there are no grammatically ill-formed elements that might interfere with the meaning, the average native or non-native speaker should eventually be able to decode the expressions and understand the underlying meaning. In this study, collocational errors of dimensional adjectives are highlighted as an example of semantic errors (see Chapter 7).

Table 3.1 The new error taxonomy according to lexical domains and sources

Source Domain	Interlingual error (L1 source)	Intralingual error (L2 source)
Semantic domain	L1-sourced semantic error	L2-sourced semantic error
Syntactic domain	L1-sourced syntactic error	L2-sourced syntactic error
Morphological domain	L1-sourced morphological error	L2-sourced morphological error
Phon/Orth domain	L1-sourced phon/orth error	L2-sourced phon/orth error

In the same vein, syntactic or morphological errors can also be affected by either L1 or L2 source. While these errors could be categorised as grammatical or morphological errors, as in the ‘linguistic category’ used by previous studies, in the current study they are subdivided by the source of the errors. For example, unaccusative verbs are a type of intransitive verbs that should not be passivised. In this study, the over-passivisation errors of non-alternating unaccusative verbs are investigated and divided by the error sources (see Chapter 8).

This study also investigates the sources of morphological errors produced by Korean learners of English. For example, it seeks to discover whether they transfer a similar L1 morphological system to the L2 production, or overgeneralise the affixation rules. In particular, the current study examines derivational morphological errors as evidence of both interlingual and intralingual influences (see Chapter 9).

It is also possible to examine phonological/orthographic errors in order to reveal their sources. Researchers have reported the interlingual effect on the English pronunciation of L2 learners (e.g. Gut, 2007; Chan & Li, 2000). Dewaele (1998) investigated a similar cross-linguistic morpho-phonological influence with Dutch learners of French and called this phenomenon ‘lexical invention’. Other researchers have investigated the orthographic errors of L2 learners that are affected by intralingual influence. For example, Laufer (1988) presented a specific type of lexical errors, termed ‘synforms’ (similar lexical forms), the phonological/orthographic/ morphological errors that are caused by similarity of forms. James (1998) reported bilinguals’ misspelling or mispronunciation errors and called these ‘misencodings’. He then divided these into interlingual and intralingual misencodings according to the sources of the errors. The current study investigates the orthographic errors made by Korean learners of English in order to show both the interlingual and intralingual influences on spelling errors (see Chapter 10).

However, as mentioned in Section 2.1.3, note that the current study exemplifies only four error features, one from each lexical domain in order to provide evidence for the grounds of the proposed model and error taxonomy, but does not aim to establish an exhaustive and comprehensive error taxonomy that includes every kind and aspect of L2 learners’ errors.

3.6 Chapter Summary

This chapter has provided the theoretical basis for a cognitive linguistic perspective to L2 learners' lexical errors. It has considered the findings of previous studies on three key issues about bilingual mental lexicon, language production models and diagnosis-based error taxonomies in preparation for the construction of a new conceptual framework to examine lexical errors and a new error taxonomy based on multi-faceted criteria.

In Section 3.1, I gave an overview of the mental lexicon, a complex language system in the human mind that is involved with vocabulary storage and retrieval. Although it is still difficult to clearly define the mental lexicon, especially for bilinguals, researchers have made efforts to elaborate various models to discover more about its nature.

In Section 3.2, I referred to three key theoretical issues explored by previous research in order to conceptualise the L2 mental lexicon: single (shared) vs. separate memory system; revised hierarchical model (RHM); language selective vs. non-selective access. Throughout the discussion on these issues, the current study assumes that L2 learners have a shared memory system for concepts linked to the semantic domain in the L2 mental lexicon and that they have separate systems for other lexical domains. This study also adopts the notion of language non-selective access, which assumes that L2 learners simultaneously activate L1 and L2 lexical information, which may cause cross-language interference in L2 production.

In Section 3.3, I introduced three main language production models in order to show the language production procedures in more detail. Levelt's (1989) L1 speech production model was a good starting point to discuss the L1 language processing, described in that model by five different components. De Bot's (2004) multilingual production model, a slightly altered version of Levelt's model, expanded the discussion from L1 to L2 language production. The

language node, added by De Bot as a new mediating component in the model, made the multilingual production model applicable to L2 language processing. Jiang's (2000) L2 vocabulary acquisition model provided a relevant conceptual framework to discuss L2 learners' lexical errors based on an L2 learners' mental lexicon that can be separated into four lexical domains. This model suggested the notion of incongruity, which could be a critical factor in producing L2 lexical errors.

In Section 3.4, I reviewed previous studies on diagnosis-based error taxonomies and showed that they had a serious limitation with regard to error categorisation, namely the somewhat ambiguous and subjective criteria used to classify errors. In order to overcome the limitation, I argued that the error taxonomy should be based on the 'source' of the errors, excluding the 'cognitive vehicles' that activate the sources in each lexical domain.

In Section 3.5, I proposed the 'L2 lexical development model', which features the four lexical domains in the L2 learner's mental lexicon, the sources of errors and the role of cognitive vehicles. I also presented an overview of the new error taxonomy based on the L2 lexical development model, which will be supported by robust empirical evidence through a corpus-based error analysis in Chapters 7 to 10. The error categories from the new error taxonomy are classified by multi-faceted criteria: the lexical domains and the sources of errors. It seems that the new error taxonomy is more sophisticated than previous diagnosis-based error taxonomies in that it removes ambiguous areas between error categories by clarifying the sources of errors and the function of cognitive vehicles. Thus, it can be argued that the error features discussed in this study will provide valuable implications for the field of language teaching and learning.

CHAPTER 4: ERROR FEATURES: SEMANTIC AND SYNTACTIC DOMAINS

As outlined in Section 3.5.2, the error taxonomy proposed in this study contains four lexical domains: semantic, syntactic, morphological and phonological/orthographic. Although there could be several kinds of errors in each domain, this study highlights one error feature from each: collocational errors of dimensional adjectives in the semantic domain, over-passivisation errors of non-alternating unaccusative verbs in the syntactic domain, derivational morphological errors in the morphological domain and spelling errors in the phonological/orthographic domain. All the errors considered here are detected in the learner corpus, the YELC (see Section 6.2.1). They can each be subdivided into two types according to their source, as interlingual or intralingual errors.

This chapter aims to provide the theoretical basis for the error features in the first two lexical domains; that is, the semantic and syntactic domains. Each is discussed within its own section, where I provide a definition, a review of the relevant literature, the significance of investigating the error feature and other important aspects to show both interlingual and intralingual influences as evidence for the grounds of the L2 lexical development model and the proposed new error taxonomy. Table 4.1 provides an overview of the chapter and introduces key concepts associated with the error features in these two lexical domains.

Table 4.1 Outline of theoretical background of the error features: semantic and syntactic domains

Domain	Error feature	Key concepts
Semantic domain	(Section 4.1) Collocational errors of dimensional adjectives	- Collocation - Dimensional adjectives - Figurative expression - Different perspectives on this type of error - How to detect the errors
Syntactic domain	(Section 4.2) Over-passivisation errors of non-alternating unaccusative verbs	- The behaviour of unaccusative verbs - Difference between alternating and non-alternating unaccusative verbs - Interlingual influence: Korean morphological system - Intralingual influence: subject animacy

4.1 Semantic Domain: Collocational Errors of Dimensional Adjectives

In the semantic domain, this study investigates collocational errors of dimensional adjectives. These errors are particularly interesting because they seem to provide important clues as to the possible sources of lexical errors in this domain from a cognitive linguistic perspective. Dimensional expressions describe semantic representations that we perceive in the three-dimensional world. They include several attributes of dimension that can be measured on a scale, for example height, length, width and weight. As the nature of dimension exists in the real world it needs to be represented in language (Dixon, 1982) and therefore most languages have dimensional expressions. These expressions are lexicalised not only in adjective forms but also in other word classes such as dimensional verbs (e.g. *expand*, *enlarge*) and dimensional nouns (e.g. *size*, *height*) (Shimotori, 2013b). However, when undertaking contrastive research, dimensional adjectives are particularly appropriate. As Shimotori (2013a: 14) indicated, “[D]imensional adjectives are well organized in their interrelationship. Prototypically a dimensional adjective correlates with another dimensional adjective with the

opposite meaning, namely its antonym, e.g. *long:short*.” Due to the nature of dimensional adjectives, we could predict that L2 learners might struggle with them in the L2 acquisition process. Indeed, the YELC (see Section 6.2.1) contains a number of collocational errors with dimensional adjectives, indicating that Korean learners of English do have difficulties in acquiring this language feature. Here is an example:

- (1) In many Internet sites, asked personal information such as real name, birthday, etc. Most of sites using personal information database, when it is hacked, the damage is very large. (Sentences from Text file 16_02 in the YELC).

In example (1), a learner is discussing the extent of the damage when hackers attack an internet site. The adjective ‘large’ does not seem appropriate as a predicate of ‘damage’. Even if this construction is altered to ‘premodifier + noun’, *large damage* still seems non-nativelike. The collocation *large damage* can be considered a deviant collocation because the adjective *large* is a very infrequent modifier for the noun *damage* according to reference corpora (i.e. BoE and BNC; see Sections 6.2.2. and 6.2.3). There may be many more collocational errors where specific adjectives are associated inappropriately with certain nouns due to semantic incongruity between L1 and L2 (see Section 2.2.1). This could be because the lexical representation of the dimensional concept is language-specific, although the representations in different language systems do sometimes overlap.

Therefore, it would be interesting to investigate whether and how these concepts are expressed in different ways in different languages. However, to date there has been very little research to investigate L2 learners’ acquisition of dimensional adjectives. The current study aims to fill this gap through a corpus-based error analysis of the L2 acquisition of

dimensional adjectives and to provide evidence of interlingual and intralingual influences on learners' errors.

4.1.1 Collocational Errors Made by L2 Learners

The dimensional adjective error is a specific kind of 'adjective + noun' combinational error. Here it is necessary to clarify the concept of 'collocation', because although the term 'collocation' is widely used, the concept is often rather ambiguous (Walker, 2008). Nation (2001: 317) simply defines the term 'collocation' as "items which frequently occur together and have some degree of semantic unpredictability". Furthermore, a number of studies that have investigated collocation have used alternative terms such as prefabricated units, phraseological units, (lexical) chunks, multi-word units and formulaic sequences. However, it seems that "the only common denominator is that the term is (at least mostly) used to refer to some kind of syntagmatic relationship between words" (Nesselhauf, 2005: 11).

According to Walker (2008), it is possible to group the different definitions of collocation into two categories: a lexical approach and a frequency (or statistically) based approach. A lexical approach uses lexical criteria such as degree of 'fixedness' or 'transparency' in order to classify different types of collocation, whereas a frequency based approach typically considers whether words co-occur more often than would be expected by chance. The current study adopts a frequency based approach, which is closely associated with the work of Sinclair (1991). More specifically, in this study the degree of 'deviation' of specific 'adjective + noun' collocations and the decision to label them as errors are determined by the frequency of the collocations in reference corpora (see Section 6.3.1).

Collocation has been a fundamental issue in the research and practice of L2 acquisition since the 1990s. Through the work of Sinclair and the COBUILD project, as Howarth (1998: 26) indicated: “A widely recognised approach to phraseology [collocation] has been familiar to ELT (English Language Teaching) professionals.” Collocational knowledge is potentially one of the most important elements in L2 acquisition and a number of researchers have investigated collocation through the analysis of learners’ errors. However, almost all studies on L2 learners’ collocational errors have focused on verb-noun collocations (e.g. Nesselhauf, 2005; Laufer & Waldman, 2011; Zheng & Xiao, 2015; Hong, Rahim, Hua & Salehuddin, 2011), presumably because L2 learners frequently make errors with verb-noun collocations. For example, Nesselhauf (2005) identified that German learners of English made more deviant verb-noun collocations with ‘lexically non-congruent’ expressions than with congruent expressions, presumably because non-congruent collocations do not have word-for-word equivalent translations between German and English, whereas congruent collocations do have such equivalents. Since this study highlights collocational errors of dimensional adjectives as one of the error types in the semantic domain, it will contribute to improving understanding of a type of adjective-noun collocational error made by L2 learners and provide empirical evidence of interlingual and intralingual influences on the production of such errors.

4.1.2 Dimensional Adjectives and the Possible Sources of the Errors

Early research into the acquisition of dimensional adjectives focused on L1 rather than L2 acquisition (Galeote & Checa, 2005; Gelman, Ravn & Maloney, 1985), as children seem to be unaware of how different dimensional adjectives are used when they acquire their first language. Clark (1973) described children's confusion with the dimensional adjectives *high* and *low* and argued that these errors are evidence of overextension of children's limited vocabulary. Even native speaker adults sometimes have difficulties with using dimensional adjectives in some contexts. For example, Maloney & Gelman (1987) suggested that native speakers could be confused by the concept of a dimensional adjective *big*. This could imply that the collocational errors of dimensional adjectives made by L2 learners might also stem from an intralingual rather than an interlingual source.

However, with regard to the L2 acquisition, previous studies have argued convincingly that one of the other main sources of collocational errors made by L2 English learners is interlingual influence (also known as L1 transfer) (Nesselhauf, 2005). Accordingly, it is reasonable to speculate that interlingual influence would also be a major factor in the specific case of collocational errors of dimensional adjectives. There have been only a few studies on the contrastive language analysis of dimensional adjectives (e.g. Kim, 2000; Kim & Wu, 2014). Shimotori (2013a, 2013b) investigated semantic similarities and differences in the conceptualisation of dimensional adjectives between Japanese and Swedish. She argued: “The basic patterns of attribution in the domain of dimension are relatively shared between languages, but it is also true that we see specific patterns of lexicalization in language” (Shimotori, 2013a: 15). That is to say, although dimensional concepts are universal, their lexicalisation might be represented in a different manner in the formation of collocations with nouns in any given language system and, consequently, the collocations of ‘dimensional

adjective + noun' can be represented differently across languages. Therefore, collocational errors of dimensional adjectives made by Korean learners of English might be related to the conceptualisation of dimensional adjectives in the Korean language.

Dimensional adjectives in the Korean language are widely used and combined with elements from different semantic domains. Kim (2011) examined eight pairs of Korean dimensional adjectives and showed that they carry prototypical meaning in the spatial domain but are also expanded to the time and abstract domains. If learners' L1 plays a significant role in English learning, then, provided that there is semantic incongruity between Korean and English, it is probable that Korean learners of English would have difficulties in the acquisition of dimensional adjectives. This is because, as mentioned above, each language has its own semantic principles of dimensional adjectives, even though the nature of dimension is universal.

In short, Korean learners of English could be subject to both interlingual influence due to the semantic incongruity and intralingual influence that also leads L1 speakers to make such errors (e.g. Clark, 1973). However, what makes the detection of collocational errors of dimensional adjectives difficult is the existence of 'figurative expressions' in languages. Since English figurative expressions can include collocations of dimensional adjectives that would otherwise appear unusual or unexpected, it can be difficult to identify which collocations are errors and which are not. This is why the current study adopts a frequency based approach to identify collocational errors of dimensional adjectives.

4.1.3 Figurative Extension and Overextension

One important semantic phenomenon we need to consider with respect to the L2 acquisition of dimensional adjectives is figurative expressions. These include, for example, metaphor, metonymy and simile, which exist in every language and culture. Such expressions are sometimes shared by different language groups, although each language has its unique figurative expressions that reflect its own culture. Dimensional adjectives often feature in figurative expressions because they have multiple meanings beyond the literal reference to the physical dimension. For example, if you look up the dimensional adjective *small* in the Merriam-Webster Online English Dictionary,⁹ the first meaning given is “having comparatively little physical size or slight dimension”. However, the extended meaning includes, for example, “lacking in strength”, “of little consequence” and “limited in degree”. In other words, the meaning of *small* is not confined to physical size but extends to somewhat abstract or figurative concepts such as ‘strength’, ‘consequence’ and ‘degree’. This polysemous nature of dimensional adjectives in figurative expressions creates a problem as to how to measure the extent of ‘deviation’ in the collocations of dimensional adjectives, because it would be difficult to draw a fine line between correct and incorrect collocations. In order to address this problem, the current study borrows from methods that are commonly used in ‘applied metaphor’ studies (see Sections 4.1.5 and 6.3.1).

As seen in the case of *small* in the Merriam-Webster Online English Dictionary, the figurative extensions of dimensional adjectives can also be considered aspects of polysemy, because, as Deignan (1999b: 319) indicated, it seems that polysemous senses of a word can be “replicated by metaphoric senses”. Metaphor is a kind of figurative extension in which lexical items are not held within the literal meaning but are expanded to abstract meanings

⁹ It can be accessed at the URL: <http://www.merriam-webster.com/dictionary>.

with the help of cognitive processes in the mental lexicon (Lakoff, 1993; Stefanowitsch, 2006). Therefore, this study will use the term ‘figurative extension’ to refer to any type of polysemous or figurative expressions, including metaphor, that are extended from literal senses of a given word but can be generally regarded as standard English expressions. Furthermore, the term ‘overextension’ of dimensional adjectives in associations with nouns will be considered as L2 learners’ errors, in contrast to the figurative extension. The overextension of dimensional adjectives may vary, because each individual L2 learner can produce their own deviant figurative overextensions (Littlemore, 2010). It is therefore important to identify these errors without any interference from the researcher’s subjectivity. Hence, the current study employs a corpus-based approach, which will be explained in Section 4.1.5.

4.1.4 Different Viewpoints Regarding Lexical Errors in the Semantic Domain

In view of the fact that, as noted in the previous section, it is at least possible to consider collocational errors of dimensional adjectives as figurative overextensions, there might be some doubt as to whether it is appropriate to call these deviant collocations ‘errors’ per se. Given that English is used throughout the world, even in countries where it is not the first language, the characteristics of ELF (English as a Lingua Franca) or World Englishes may well be diverse, because the output of language always interconnects with speakers’ first languages and their own culture and circumstances (Crystal, 2003; Seidlhofer, 2005). From the perspective of ELF, as Kachru (1985) noted, the spread of English around the world is not restricted to the Inner Circle of countries where English is the primary medium of communication, such as the USA, the UK, Canada and New Zealand, but extends to the

Outer Circle of areas in which English is used as a second language (ESL) and the Expanding Circle where English is taught as a foreign language (EFL), as in Korea. This may mean that even if some expressions deviate from so-called ‘standard English’ used in the Inner Circle, they should not necessarily be labelled ‘errors’. Rather, they might be accepted by the speakers of the same first language and culture group as expressions that are probably unique, but still understandable.

Similarly, as mentioned in Section 4.1.3, figurative overextensions could be construed as valid expressions from the perspective of ELF. They are, however, potentially more diverse than conventional expressions in ELF because they often include creative expressions produced by individuals in the different contexts.

Another linguistic phenomenon in the semantic domain is the use of communication strategies, as mentioned in Section 3.4. The term ‘communication strategy’ appears in the literature of L2 acquisition and is generally defined as “the speaker’s attempt to communicate meaningful content in the face of some apparent deficiencies in the interlanguage system” (Tarone, 1981: 286). According to Tarone (1981), when L2 learners attempt to communicate thoughts that their interlanguage structures are inadequate to convey, they activate these communication strategies. Similarly, Varadi (1983) focused on the situation in which L2 learners have gaps between their interlanguage and the target language, to which he gave the term ‘hiatus’. He described communication strategies as “the strategies the learner employs when he experiences a ‘hiatus’ in his interlanguage repertoire” (Varadi, 1983: 80). These strategies include various kinds of cognitive processes, such as paraphrase, borrowing and avoidance.

Interestingly, the deviant expressions produced via the three pathways explained above, i.e. language variation (ELF), creative use (figurative expressions) and communication strategy, seem to be activated by the cognitive vehicles, which this study claims encompass the general mental (cognitive) processes (see Section 3.4.2). For example, cognitive vehicles might operate more dynamically when figurative expressions are created, which would increase the likelihood that certain language users would frequently create deviant expressions. In other words, the degree of deviation of these expressions could be determined by how far the cognitive vehicles extend the lexical information in the mental lexicon.

Notwithstanding the issue as to whether these deviant expressions might be considered as different kinds of unique but acceptable expressions, from the perspective of language teaching and learning this study argues that they are, indeed, ‘errors’. As James (1998: 45) noted, no L2 learner would be willing to learn, for instance, ‘Brazilian English’, rather than American or British English. It would also be foolish if L2 learners were forced to learn deviant expressions, even if these were not considered ‘errors’. L2 learners will wish to correct their lexical errors if they realise that the expressions sound strange to native speakers, even though they are grammatically correct. In the same vein, English teachers will certainly wish to correct the deviant expressions so that students do not use somewhat strange expressions again, setting aside the question of how to correct such errors. In other words, from the ELT point of view, these deviant expressions should be dealt with in the same way that other kinds of errors (e.g. grammatical errors) are corrected. This standpoint might raise the question of how to define so-called ‘standard English’ as a norm to be distinguished from the deviant expressions, a question that itself could raise some sensitive political issues (James, 1998). However, that question is not within the scope of the current study, although it would be an important topic for future investigation.

Still, it might be very difficult to draw a fine line between ‘errors’ and correct usage, especially in the semantic domain. That is why the current study adopts a corpus-based analysis to identify collocational errors of dimensional adjectives and the sources of the errors, unlike previous studies in which researchers used their L1 intuition to determine the lexical errors.

4.1.5 Corpus-based Approach to Detect Collocational Errors of Dimensional Adjectives

As mentioned above, this study adopts a corpus-based approach to detect collocational errors of dimensional adjectives and these errors can be regarded as figurative overextensions. The main problem in the detection of figurative overextensions is how to set up a clear criterion between acceptable and unacceptable figurative expressions made by L2 learners. Note that figurative extensions are also conventionalised to the degree that they become part of the native language (Steen, 2007). Hence, we need a relevant norm that can reflect contemporary English. In this study two reference corpora, the BoE and the BNC, were used to aid the determination of these types of errors (see Section 6.3.1).

It is important to note that collocational errors of dimensional adjectives, like other semantic errors, should be perceived as being on a continuum of deviation, rather than in terms of a dichotomous labelling of errors. This is one of the main reasons why reference corpora are used as a norm to detect the degree of figurative overextension. The same corpus-based method has been employed in other studies to identify figurative expressions (Deignan, 1999a; Li, 2015). As Deignan (1999a: 178) indicated: “A computerised corpus can enable the researcher to detect patterns of usage more quickly than either the use of intuition or the

analysis of individual texts, as words or expressions are automatically retrieved from the corpus and sorted. This can also, arguably, lead to a less subjective analysis.”

The corpus-based method used by the current study to detect figurative overextensions is similar to the method that is widely used in applied metaphor studies. For researchers of applied metaphor, identifying metaphoric expressions has presented a major challenge owing to the problem of the degree of ‘metaphoricity’, whereby “some metaphors are more metaphorical than others” (Cameron, 1999: 107). In that case, the main problem is how to establish a relevant set of criteria that can be used to detect whether a particular word or sequence is being used metaphorically. Hence, it is not surprising that researchers have tried to establish a reliable method to identify metaphoric expressions in real discourse. One seminal method is the Metaphor Identification Procedure (MIP) proposed by the Pragglejaz Group (2007), who explained that “the MIP was designed to correctly discriminate metaphor from other types of meanings” (Pragglejaz Group, 2007: 31). The MIP includes a series of procedures in which several experts of metaphor check every single lexical unit in context to see whether or not they are metaphoric; the results are then reviewed through an inter-rater reliability test. Since the MIP was introduced, other types of methods to identify metaphors have been developed (Steen, 2007). In the current study, the procedures in which native speakers identify the degree of deviation are substituted by corpus-based analysis. The figurative overextensions are identified through comparison with the reference corpora (see Section 6.3.1 for more information about research procedures). There has been very little research on the ability of L2 learners to use figurative expressions (Littlemore & Low, 2006) and consequently not many attempts to detect figurative overextensions made by L2 learners. Therefore, the current study aims to improve the understanding of deviant figurative expressions (or overextensions) made by L2 learners (see Chapter 7).

4.2 Syntactic Domain: Over-passivisation Errors of Unaccusative Verbs

With regard to the syntactic domain, the current study examines a particular type of grammatical error: the over-passivisation of English unaccusative verbs. Because English unaccusative constructions are difficult to acquire, though not impossible, the L2 acquisition of the unaccusatives has attracted the interest of linguists and language teachers. Indeed, several previous studies have examined these kinds of errors produced by L2 learners with various L1 backgrounds (Balcom, 1997; Chung, 2014, 2015; Kim, 2010; Mo, 2015; Oshita, 1997; Zobl, 1989; Dolgormaa & Lee, 2011). However, the findings of previous studies have reached sometimes conflicting conclusions regarding the possible causes of the over-passivisation errors.

Unaccusative verbs are a type of intransitive verbs that cannot be passivised. However, L2 learners of various L1 backgrounds frequently produce inappropriate passive errors, as illustrated in (2).

- (2) a. *The world war III will be happened. [Chinese] (Yip, 1995: 129)¹⁰
b. *The strange event was occurred last May. [Unknown] (Hubbard & Hix, 1988: 94)
c. *I was nearly arrived to my office. [Italian] (Oshita, 2000: 314)
d. *He is also appeared on the list of investigations of gold smuggling. [Spanish] (Oshita, 2000: 314)
e. *My mother was died when I was just a baby. [Thai] (Zobl, 1989: 204)
f. *First, the change of life-style will be happened. [Korean] (Ju, 1997, cited from No & Chung, 2006)

L2 learners experience difficulties in the acquisition of English unaccusative verbs, probably because of their unique nature. Intransitive verbs in English can be divided into two

¹⁰ The asterisk (*) is tagged in the ungrammatical structure and the word followed by the angle bracket (<) is the target form of the ungrammatical item.

subclasses: unergatives and unaccusatives (Burzio, 1986; Perlmutter, 1978). The unergatives (e.g. *swim*, *jump*) have a grammatical subject with the thematic role of Agent, which follows a canonical linking rule between the subject and the verb, whereas the unaccusatives (e.g. *disappear*, *happen*) adopt a thematic role Theme as their subject, but a Theme is canonically linked to an object position in a sentence, not a subject. It is this asymmetry that seems to cause the errors of unaccusative verbs as in (2) and this is probably one of the reasons why the unaccusatives have been observed and investigated by many L2 researchers and teachers.

The unaccusatives again comprise two subclasses: alternating and non-alternating.¹¹ The alternating unaccusative verbs have transitive counterparts so that they can be used as active forms as well as passive forms, as in (3). In contrast, the non-alternating unaccusative verbs do not have corresponding transitive counterparts. Hence, they do not take an object in a sentence, which means they cannot be passivised, as in (4a) and (4c). Accordingly, the three sentences with the alternating verb *break* in (3) are all grammatically correct, whereas in the case of the non-alternating verb *arrive*, only sentence (4b) in (4) is grammatical, but (4a) and (4c) are not. The classification of English intransitives can be easily summarised as in Figure 4.1.

- (3) a. The boy broke the window.
b. The window broke.
c. The window was broken. (a-c from Chung, 2014)

- (4) a. *They arrived the guest yesterday.
b. The guest arrived yesterday.
c. *The guest was arrived yesterday. (a-c from Chung, 2014)

¹¹ Instead of ‘alternating’ and ‘non-alternating’, some scholars (e.g. Yip, 1995) have used the terms ‘paired’ and ‘unpaired’ respectively.

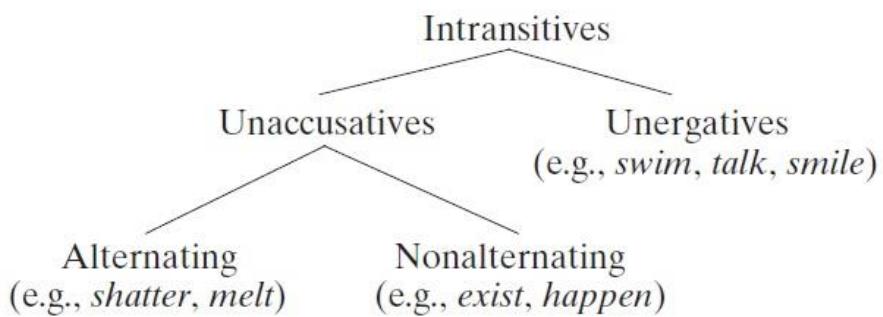


Figure 4.1 The classification of English intransitive verbs (Oshita, 2000: 295)

4.2.1 The Possible Causes of Over-passivisation Errors of Unaccusative Verbs

Many researchers have explained the unaccusative errors from the perspective of Generative Grammar. They have regarded these phenomena basically as a nominal phrase (NP) movement in the D-structure (Balcom, 1997; Oshita, 1997, 2000; Zobl, 1989), although they have given somewhat differing accounts of the errors. For example, Zobl (1989) claimed that learners use unaccusatives under the syntactic rule for passive formation and Balcom (1997) supported Zobl's account. Oshita (2000) argued that the over-passivisation of English unaccusatives is due to learners' overgeneralisation of the passive morphosyntax of the target English.

In contrast, the current study does not view over-passivisation errors of unaccusative verbs from the perspective of Generative Grammar. Instead, it tries to find the possible causes of the errors in order to provide pedagogic implications for ELT. Previous studies have cited a number of possible causes of passivisation errors of English unaccusatives, but these can be narrowed down as follows, depending on which features the researchers highlighted:

- 1) A causer effect (animacy or pragmatic view) (e.g. Ju, 2000; Pae, Schanding, Kwon & Lee, 2014)

- 2) L1 influence (e.g. Lee, 2009; Montrul, 1999; Park, 2005)
- 3) L2 input (e.g. Ahn, 2015; Chung, 2015; Han, 2006; Hwang, 2006; Kim, 2004)
- 4) Multiple factors (e.g. Chung, 2014; Lee & Choi, 2011; No & Chung, 2006)

First, Ju (2000) proposed a conceptualisable agent as a possible cause for the over-passivisation of unaccusatives. She suggested that a causer (external or internal) of an event in the discourse tends to make learners produce a passive form, as in (5).

- (5) a. A fighter jet shot at the ship. The ship sank slowly.
- b. The rusty old ship started breaking up. The ship sank slowly.

(a-b from Ju, 2000: 92)

Comparing (5a) with (5b), one can note that the causer of the event ‘the ship sank’ in (5a) is a fighter jet and thus the event is externally caused, whereas in (5b) there is no overt causer of the event, which therefore is internally caused. Ju (2000) found that learners tend to over-passivise unaccusatives with an external rather than internal causer. Similarly, Pae et al. (2014) focused on a subject animacy effect in over-passivisation errors and found that L2 learners tend to make more of these errors with inanimate subjects.¹²

Secondly, Montrul (1999) claimed that L1 influence plays a significant role in the L2 acquisition of alternating unaccusative verbs. Montrul investigated the L2 Spanish acquisition of L1 English speakers and L1 Turkish speakers and found that L1 English speakers had difficulty with Spanish alternating verbs, whereas L1 Turkish speakers did not. Considering that English has zero morphology for alternation while Turkish, like Spanish, has overt

¹² This result from Pae et al. (2014) is in conflict with the findings of the current study as to subject animacy effect, which will be discussed in more detail in Chapter 8.

morphology, the results suggested that the interlingual influence (L1 transfer) plays a significant role in the L2 acquisition of the alternating unaccusative verbs.

Thirdly, just as for other features of language acquisition, there are various accounts of L2 input. Several scholars have examined the acquisition of unaccusative verbs according to proficiency levels or age groups. For example, Han (2006) reported that the L2 acquisition of unaccusative verbs could become fossilised, probably because of the lack of L2 input. This suggests that L2 input or instruction seems to be an important factor, a conclusion reinforced by other studies (e.g. Ahn, 2015; Chung, 2015; Hwang, 2006; Kim, 2004).

Finally, some studies have tried to show how multiple factors including the three above affect the L2 acquisition of unaccusative verbs, where the degree of influence of each factor differs among studies (e.g. Chung, 2014; Lee & Choi, 2011; No & Chung, 2006). For example, No & Chung (2006) have examined how each of the three factors (i.e. English inherent factors, L1 influence and semantic factors) affected the L2 Korean students' acquisition of unaccusative verbs.

The current study investigates the possible sources of errors in the L2 acquisition of English non-alternating unaccusative verbs as evidence of L2 learners' errors in the syntactic domain based on the proposed new error taxonomy. It shows that both interlingual and intralingual influences play a significant role in the L2 acquisition of non-alternating unaccusative verbs. With regard to the interlingual influence, it looks at whether L1 morphology affects over-passivisation errors. For the intralingual influence, it examines the subject animacy effect on the errors. Using corpus-based error analysis, this study explores the errors of specific English non-alternating unaccusative verbs in a learner corpus that consists of free compositions written by university students in Korea (see Section 6.2.1). It also investigates

whether the distribution of errors of non-alternating unaccusative verbs varies according to proficiency level, in order to see how L1 morphology and subject animacy affect the L2 acquisition throughout the development of competence (see Chapter 8).

4.2.2 The Interlingual Influence: L1 Morphology

Some researchers have conducted cross-linguistic investigations into the morphological factor that affects the L2 acquisition of unaccusative verbs. The L2 acquisition could be influenced by whether or not the learners' L1 has a similar morphological system of unaccusative verbs. For instance, it is recognised that English has zero morphology regarding unaccusativity, as in (6), where the verb form *break* is exactly the same regardless of whether it is being used as a transitive or intransitive verb. On the other hand, some languages, such as Spanish, Turkish, Japanese and Korean, use particular morphemes to distinguish the alternation of unaccusative verbs.

- (6) a. The man broke the window.
b. The window broke. (a-b from Montrul, 2000: 234)

Montrul (1999) used a picture judgement task with 54 subjects split into two experimental groups, one comprising English-speaking learners of Spanish and one made up of Turkish-speaking learners, along with a control group of Spanish native speakers. Spanish and Turkish both have overt morphology of alternating unaccusative verbs, whereas English has zero morphology in verb alternation. The study showed that L1 Turkish learners of Spanish made fewer errors with Spanish alternating unaccusative verbs than L1 English learners of

Spanish did. Given that Spanish and Turkish alternating unaccusative verbs behave alike, the findings suggest that the L1 has a significant influence on the L2 acquisition of Spanish alternating unaccusative verbs.

Similarly, Kondo (2005) conducted an experiment on the use of unaccusative verbs by Japanese and Spanish learners of English. As mentioned, Spanish has overt morphology regarding verb alternation, while English has zero morphology. Japanese also has causative and anti-causative morphology (Montrul, 2001); however, Japanese morphological marking of unaccusative verbs is more complicated than Spanish in that there are two different types of morphemes, which are either lexically or syntactically involved in causative constructions. Accordingly, the results of the study revealed that Spanish learners of English were significantly more likely to over-passivise alternating unaccusative verbs (e.g. *close*, *freeze*, *dry*) than non-alternating unaccusative verbs (e.g. *die*, *appear*), whereas Japanese learners of English tended to over-passivise both alternating and non-alternating verbs in equal measure. In other words, Spanish learners make fewer errors with non-alternating unaccusative verbs than Japanese learners and this difference might derive from the complexity of Japanese morphology.

Chung (2014) investigated the influence of L1 morphology among Korean and Chinese learners of English. Unlike Korean, but in common with English, Chinese does not have any morphological markers for verb alternation (Ju, 2000). Therefore, we would expect that Chinese learners of English would make fewer errors of unaccusative verbs than Korean learners of English, based on the L1 morphological factor. Chung (2014) compared the results of grammaticality judgement tasks between English alternating and non-alternating

unaccusative verbs for Korean ($N=117^{13}$) and Chinese ($N=99$) college students and high school graduates and found that L1 Korean subjects made significantly more errors with alternating unaccusative verbs than with non-alternating unaccusative verbs. In contrast, for L1 Chinese subjects there were no significant differences in the results for the two types of verbs. The results suggest that the L1 morphological factor does play a role in the acquisition of these verbs. Moreover, according to the study, L1 morphological influence seems to exert a greater effect on the acquisition than other factors such as discourse (external causation) or animacy.

However, complexity of the L1 morphological system might not always cause difficulties with unaccusative verbs. According to Hwang (2006), in the Japanese language passive forms occur more frequently than in the English or the Korean language. Therefore, one could expect that if the L1 has an influence on the acquisition process, this might be an obstacle to the L2 acquisition of English unaccusative constructions by Japanese learners. Yet the frequent use of passive constructions does not necessarily mean that Japanese learners of English would be likely to over-passivise English unaccusatives and have more difficulty in mastering English unaccusative constructions. Contrary to Hwang's (2006) prediction, his investigation of the acquisition of unaccusative verbs by Korean and Japanese learners of English found that in grammaticality judgement tasks, Japanese subjects scored significantly more highly than Korean subjects. The results were the exact opposite of Hwang's prediction that Korean subjects would be better than Japanese subjects in the L2 acquisition of unaccusative verbs.

The above review of previous studies reveals conflicting results concerning the interlingual influence on unaccusative constructions. This gives rise to some questions: Based on a

¹³ The capital N means the number of subjects (sample size).

sample of naturally occurring data, does L1 morphology play a role in the L2 acquisition of unaccusative verbs, and if so, to what extent? If it does play a role, then why have previous studies reached different conclusions? The major difference between the current study and previous research is that this study is based on corpus-based analysis that deals with an exceptionally large and naturally occurring dataset (see Sections 2.3.2 and 6.3.2), whereas previous studies examined data from grammaticality judgement tasks with relatively small datasets.

4.2.3 Korean Morphological System of Passives and Causatives

Korean is one of the agglutinative languages, which contain abundant morphemes. Korean words link various morphemes to create causative/anti-causative meanings and this can present difficulties for foreign learners of Korean (Kim, 1996). It is well-known that the passive construction is one of the most difficult aspects of English-Korean translation and linguists have explained that this is because Korean tends to use a topic marker *-un* or *-nun* to avoid passive constructions (Cho, 2005). In addition, Korean passive construction is very similar to causative construction and they even share the same morphemes (e.g. *-i-*, *-hi-*, *-li-*, *-ki-*). Therefore, even Korean native speakers are sometimes not aware of the difference between them (Kim, 1998).

There are three kinds of passivisation in the Korean language: lexical, morphological and syntactic, used according to the characteristic of each verb (Lee, 1990). First, lexical passivisation refers to the use of verbs that have passive meaning. Montrul (1999: 193) presented this kind of passive in Spanish, which she termed ‘a lexically suppletive causative counterpart’ (e.g. *die-kill*). Second, morphological passivisation is probably the most typical

way to passivise verbs in Korean, although it is not applicable to every verb. Third, the so-called ‘syntactic passive’ is used to passivise verbs that do not have a relevant passive morpheme or a lexically suppletive counterpart. To make a syntactic passive construction, the auxiliary verb *-cita* is attached to the verb, along with the morpheme *-u-*. The current study focuses particularly on the morphological passive.

The asymmetry of passive forms between English and Korean equivalents may cause difficulties in the L2 acquisition of unaccusative verbs. While some English non-alternating unaccusative verbs have Korean equivalents, others do not, and it is to be expected that Korean learners of English will have particular difficulty with the latter type. Therefore, this study chooses six non-alternating unaccusative verbs: three ‘matched’ and three ‘mismatched’ verbs (see Section 6.3.2). The comparison between these two kinds of verbs will show the interlingual influence in the L2 acquisition of English unaccusative verbs (see Chapter 8).

It is worth noting that the current study deals only with non-alternating unaccusative verbs. It does not consider alternating unaccusative verbs, because it would be very difficult to judge the grammaticality of alternating unaccusative verbs in learner corpus data. In order to determine the grammaticality of passive uses of alternating unaccusative verbs one would need to find out the writers’ intentions, because these verbs can also be passivised. Given the nature of the data used here that would be an extremely difficult, if not impossible, task. On the other hand, because non-alternating unaccusative verbs cannot be passivised, it is easy to determine whether their use is ungrammatical.

4.2.4 The Intralingual Influence: Subject Animacy

Animacy of subject nouns has been considered an important factor that affects syntactic and semantic features of verbs (e.g. verb transitivity) (Hinkel, 2002). It has been reported that animacy has an influence on the L1 acquisition of English transitive verbs (Dewart, 1979; Lempert, 1990). For example, Dewart (1979) demonstrated that children had a strong preference for animate subjects with passivised verbs when the subjects were ‘acted-upon’ elements. Prat-Sala, Shillcock & Sorace (2000: 112) also showed that there is “an influence of animacy upon syntactic process in spoken production by Catalan-speaking children”. The Catalan-speaking participants tended to produce a certain type of word order more frequently when the thematic role Patient (or Theme) was animate than when it was inanimate. This can be explained by the fact that an animate subject is canonically linked to a subject position because animacy is a typical property of Agent, whereas an inanimate subject is linked to an object position because inanimacy is a property of Theme. However, non-alternating unaccusative verbs in English adopt Theme, which typically has an inanimate thematic property, in a subject position. Consequently, if an inanimate subject is used with a non-alternating unaccusative verb, L2 learners would tend to incorrectly passivise the verb.

In their investigation into the animacy effect on the L2 acquisition of unaccusative verbs, Pae et al. (2014) administered a computerised grammaticality judgement task and showed that Korean-speaking learners of English tended to make significantly more over-passivisation errors with unaccusative verbs than English native speakers did. Pae et al. argued that transitivity is closely related to the subject noun’s animacy and volitionality. Therefore, when L2 learners use unaccusative verbs, they might get confused and assume that the subject takes the thematic role of Agent. Chung (2014) also looked into the subject animacy effect of unaccusative verbs. His experiment showed that Korean learners of English made

significantly more errors with inanimate subjects than with animate subjects. This might be because the learners assumed that the subjects of unaccusative verbs took the thematic role of Agent, even though they actually had the property of Theme, which resulted in over-passivisation errors of the unaccusative verbs. The current study argues that the subject animacy effect with unaccusative verbs is undoubtedly a characteristic of the target language (English), but not of the learners' L1 (Korean). This is to say, the over-passivisation errors caused by the subject animacy effect appear to derive from the peculiar role assignment of English unaccusative verbs, not from an attribute of the Korean language, because unaccusative verbs do not exist in Korean. Consequently, the over-passivisation errors that result from the subject animacy effect would seem to be due to intralingual influence because they are broadly related to the incorrect application of passivisation rules of English unaccusative verbs.

In summary, this research provides evidence of both interlingual influence (L1 morphology) (see Section 4.2.3) and intralingual influence (subject animacy effect) in the L2 acquisition of non-alternating unaccusative verbs. As was explained with regard to the interlingual influence, this study differs from previous research chiefly in the data it employs. Specifically, this study analyses L2 learners' naturally occurring data (from a learner corpus) (see Section 6.2.1), rather than data from grammaticality judgement tasks. Therefore, it will contribute to the knowledge of differences between learners' receptive and productive language skills (see Chapter 8).

4.3 Chapter Summary

This chapter has aimed to provide a theoretical basis for the error types from the first two of the four lexical domains, according to the new error taxonomy proposed in this study.

In Section 4.1, I introduced the dimensional adjectives and the concept of collocation used in this study. By investigating the collocational errors of dimensional adjectives, I highlighted the existence of semantic incongruity in the semantic domain and showed that both interlingual and intralingual influences could play a role in producing incorrect collocations with regard to dimensional adjectives.

In Section 4.2, in order to show the interlingual and intralingual influence in the syntactic domain, I focused on English unaccusative verbs. These verbs behave in a unique way and seem to present difficulties for many L2 learners from various L1 backgrounds. I suggested that the L2 acquisition of non-alternating unaccusative verbs is affected by interlingual influence through the L1 morphological effect and by intralingual influence through the subject animacy effect.

The next chapter will provide the theoretical background to the other two error features considered in this research, which are from the morphological and phonological/orthographic lexical domains.

CHAPTER 5: ERROR FEATURES: MORPHOLOGICAL AND PHONOLOGICAL/ORTHOGRAPHIC DOMAINS

This chapter provides the theoretical basis for the error features in the morphological and phonological/orthographic domains. In the same way as Chapter 4, which gave the theoretical background to the error features in the semantic and syntactic domains, it presents a review of the previous research and discusses the significance of the corpus-based error analysis with these error features, along with other important matters related to both interlingual and intralingual influences. Table 5.1 provides an overview of the chapter with key concepts in the two lexical domains.

Table 5.1 Outline of theoretical background of the error features: morphological and phonological/orthographic domains

Domain	Error feature	Key concepts
Morphological domain	(Section 5.1) Derivational morphological errors (distributional knowledge)	- Morphological awareness - Three aspects of derivational morphology - Distributional knowledge of derivational morphology
Phonological /Orthographic domain	(Section 5.2) Orthographic (spelling) errors	- Why spelling errors have been neglected - Interlingual influence: L1 phonological and orthographic interference - Comparison between English and Korean consonant systems - Intralingual influence in inflection and derivation

5.1 Morphological Domain: Derivational Morphological Errors

In the morphological domain, this study examines the derivational morphological errors in English in order to show evidence of both interlingual and intralingual influences in the L2 acquisition of morphology. It highlights in particular the aspect of distributional knowledge, which is one of the key elements to measure morphological awareness (Kuo & Anderson, 2006) (see Section 5.1.2). This section covers the concepts of morphological awareness, three aspects of derivational morphological knowledge and the importance of distributional knowledge and productive learner data. In order to provide a theoretical background for this study it also reviews previous research into the acquisition of derivational morphology.

5.1.1 Metalinguistic Awareness and Morphological Awareness

Metalinguistic awareness, which refers to the ability to reflect on and manipulate different linguistic units (Adams, 1994; Perfetti, 2003), has been recognised as an important ability that is necessary for language acquisition (Nagy, 2007). It includes phonological, orthographic and morphological awareness of language forms. Morphological awareness, which can be simply defined as “a learner’s grasp of morphological structure as well as his or her capability of using this knowledge” (Koda, 2000: 299), has attracted researchers’ attention because of its strong correlation with learners’ vocabulary knowledge. As Hayashi & Murphy (2011: 105) indicated: “Developing morphological awareness is an essential component of vocabulary growth, given that it can contribute to enhanced depth of vocabulary knowledge and provide a pathway to deeper associations with more members of a word family.” In other words, as learners’ morphological awareness, such as knowledge of affixation, develops, the size of the vocabulary they acquire increases (Anglin, 1993; Bauer &

Nation, 1993). Consequently, it is not surprising that a number of studies have investigated the effects and the sources of morphological awareness with regard to first and second language acquisition (Carlisle, 2000; Choi, 2015; Kan, 2014; Lowie, 2000; Hayashi & Murphy, 2011).

Morphological awareness can be divided into three components: awareness of inflectional, compounding and derivational morphology (Kuo & Anderson, 2006). Inflectional morphology usually serves a grammatical function such as marking number (e.g. plural -(e)s), or tense (past tense morpheme -(e)d, third person present tense -(e)s) and is regarded as the easiest morphological system for learners to acquire, in that it does not alter the syntactic category of the word (Bauer & Nation, 1993; Katamba & Stonham, 2006). Compounding morphology is a very important way of creating new words by means of combining at least two bases, called root morphemes. The words made by the process of compounding are relatively rare in English and the root morphemes used are generally too arbitrary for learners to generalise the affixation rules. Awareness of compounding morphology is therefore not considered a key element of morphological awareness except in certain circumstances. However, awareness of derivational morphology has generally been recognised as a significant factor in L2 acquisition, because learners may be able to expand their vocabulary by applying the affixation rules (Kuo & Anderson, 2006). Due to this generalisability and adaptability of knowledge of the derivational aspect of morphology, L2 learners sometimes make errors in producing vocabulary where they overgeneralise the affixation rules.

The majority of previous research on morphological awareness has analysed data from reading comprehension tasks or written elicitation tasks. However, some studies have dealt with L2 learners' errors drawn from naturally occurring data. For example, James (1998) presented some derivational errors as a category of 'lexis errors' as in (1). Al-Shormani & Al-

Sohbani (2012) also showed morphological errors made by Yemeni university students. They categorised the error types as ‘mis-selection of prefix’ (e.g. (2a), (2b)) and ‘mis-selection of suffix’ (e.g. (2c), (2d)). Rajab, Darus & Aladdin (2016) illustrated similar errors made by Libyan postgraduate students as in (3).

- (1) a *colorfuller (<more? colorful) scene
b. visit me *soonly (<soon)

(a-b from James, 1998: 154)

- (2) a. He is *unsane (<insane)
b. I am *nonhappy (<unhappy)
c. Dr. Mahmoud’s *honestness (<honesty)
d. I am an *ambitionable (<ambitious) person in my life

(a-d from Al-Shormani & Al-Sohbani, 2012: 123)

- (3) to increase the knowledge and referashing the *thinkfull (<thinking) [sic]
(from Rajab et al., 2016: 284)

However, while the above mentioned studies reported the learners’ errors based on the researchers’ own error taxonomies, they did not discuss the sources or factors that influence the morphological errors or the relationship between morphological awareness and learners’ writing proficiency. Almost all the other studies examined how morphological awareness affected learners’ reading comprehension (e.g. Carlisle, 2000; Choi, 2015; Jeon, 2011; Wang, Cheng & Chen, 2006) or vocabulary acquisition (e.g. Bae, 2015; Choi, 2015; Yilmaz, 2014; Hayashi & Murphy, 2011; Wang et al., 2006; Ramirez, Chen, Geva & Kiefer, 2010). To date,

only a few studies have investigated the contribution of morphological awareness to writing proficiency, taking as their informants either L1 children in the early years of elementary school (e.g. Apel & Werfel, 2014) or L2 secondary school students aged from fifteen to eighteen (e.g. Leontjev, Huhta & Mäntylä, 2016). Therefore, the current research, which investigates the sources of derivational morphological errors with learner corpus data, is expected to fill the gap with regard to the L2 acquisition of derivational morphology.

5.1.2 Three Aspects of Derivational Morphology

Tyler & Nagy (1989) noted that knowledge of derivational morphology comprises three aspects: relational, syntactic and distributional knowledge. They argued that relational knowledge involves “recognizing that words have complex internal structure and that two or more words may share a common morpheme” (Tyler & Nagy, 1989: 249). In other words, it is the ability to see morphological relationships between two or more words that share a common base morpheme. For example, learners with good relational knowledge will know that the word *player* can be split up into the verb *play* and the suffix *-er* (Kuo & Anderson, 2006). Syntactic knowledge is “knowing that derivational suffixes mark words for syntactic category” (Tyler & Nagy, 1989: 249). For example, learners can figure out that the word *satisfaction* derived from the verb *satisfy* is a noun and that the word *satisfactory* is an adjective. If they reach their conclusions based on the suffixes (i.e. *-tion* from *satisfaction* and *-ory* from *satisfactory*), then we can say that they have syntactic knowledge of the given suffixes. Finally, distributional knowledge refers to “the understanding of how affixes are constrained by the syntactic category of the stems they attach to” (Kuo & Anderson, 2006: 166). Lardiere (2006: 73) referred to distributional knowledge using the term ‘selectional

knowledge', because "this is the knowledge of selectional restrictions on the concatenation of stems and affixes". For example, it is the knowledge that the suffix *-ness* can be attached to an adjective *quiet* to make *quietness*, but not to a verb *play* to make *playness*.

Among the three aspects of knowledge of derivational morphology, it appears that the most difficult and the last to be acquired is distributional knowledge. According to Kuo & Anderson (2006: 167), this is because "without being able to recognise the stem in a complex word or differentiate different syntactic categories, one can hardly see the distributional constraints". It might also be due to the fact that distributional knowledge should be gauged only by productive language skills, whereas relational and syntactic knowledge can be easily measured by receptive skills, particularly reading. For example, relational knowledge is often assessed by a word segmentation task, where the informants are presented with complex words and then asked to segment them. When testing syntactic knowledge, researchers often employ a grammaticality judgement task. Accordingly, the majority of previous studies related to morphological awareness have focused on either relational or syntactic knowledge of derivational morphology rather than distributional knowledge (e.g. Carlisle, 2000; Kieffer & Lesaux, 2012; McBride-Chang et al., 2005). In contrast, in order to investigate distributional knowledge, one would need appropriate productive learner data. Given that, as mentioned above, distributional knowledge is more difficult and acquired later than relational and syntactic knowledge, research into this aspect using L2 learners' naturally occurring data could provide valuable evidence of the effect of morphological awareness. Indeed, there have been some studies that analysed learners' productive data. However, these studies used elicited vocabulary production tasks in a predetermined test setting (Carlisle, 2000). That is, the experimental tasks were a somewhat indirect way of obtaining productive data and those data were quite different from the naturally produced learner corpus data used in the current

study (see Section 6.2.1). Since the current study deals with appropriate and productive learner data, it is expected that the findings will provide a more accurate reflection of the L2 acquisition of derivational morphology.

5.1.3 The Effect of Morphological Awareness on L2 Acquisition

It has been largely accepted that morphological awareness has a ‘facilitative effect’ (Schmitt & Zimmerman, 2002), which implies that once language learners know a base word or a derived word they can easily recognise the members of the word family (Bauer & Nation, 1993). The contribution of the facilitative effect of morphological awareness to reading comprehension and vocabulary acquisition has been particularly well documented. As Kuo & Anderson (2006: 171) indicated: “Awareness of derivational morphology is perhaps the most widely studied aspect of morphological awareness in reading research and is usually considered to be a general indicator of morphological awareness.” Carlisle (2000) investigated the effect of morphological awareness on reading comprehension among third and fifth grade English native students. McBride-Chang et al. (2005) carried out a cross-linguistic study with Chinese and Korean learners and found that morphological awareness contributes to reading ability. Jeon (2011) conducted an experiment with Korean tenth grade students and showed that, among six reading and language-related variables, derivational morphological knowledge appeared to be particularly important in L2 reading comprehension. In addition, there have been some intervention studies that have demonstrated the effects of morphological awareness instruction on word acquisition or reading comprehension (e.g. Kan, 2014; Goodwin & Ahn, 2015; Li & Chen, 2016). In general, therefore, a review of previous studies shows that derivational morphology is facilitative particularly to reading (Schmitt &

Zimmerman, 2002). As Carlisle (2000: 170) claimed: “Morphological awareness, as it contributes to reading, must have as its basis the ability to parse words and analyse constituent morphemes for the purpose of constructing meaning.”

On the other hand, there has been little research into morphological awareness using L2 learners’ productive data. This is probably because productive language skills are more complex to examine than receptive skills (Schmitt & Zimmerman, 2002). More importantly, as mentioned in the previous section, it is very hard to gain relevant productive data for such research (Carlisle, 1996; Green et al., 2003). Only recently has attention been paid to the role of morphological awareness in writing skills. Leontjev et al. (2016) investigated the correlation between word derivational knowledge and writing proficiency of adolescent L2 learners of English. They showed that students’ derivational knowledge is strongly correlated to their essay writing ability. McCutchen & Stull (2015) also examined how children’s morphological awareness relates to morphological derivations in their writing, although the children were fifth grade English native speakers, not L2 learners. Interestingly, children made specific derivational errors, newly created deviant derivatives that McCutchen & Stull referred to as ‘morphological invention’. Such inventions can be regarded as derivational errors because they do not exist in English. The frequency of morphological invention was proportional to the degree of morphological awareness. In other words, children with higher morphological awareness made more morphological inventions. McCutchen & Stull’s study shows that the degree of morphological awareness in language production could be considerably different from that in comprehension. Hayashi & Murphy (2011) also suggested that productive morphological awareness is cognitively more demanding than receptive morphological knowledge. Similarly, Schmitt & Meara (1997) used pre- and post-tests of

word association and suffix knowledge with Japanese learners of English and found that the learners had better receptive knowledge than productive knowledge.

The phenomenon of morphological invention highlighted by McCutchen & Stull's (2015) study might offer a clue to reveal the cognitive processes of L2 learners' language production with regard to morphological awareness. It is presumably a by-product of the overgeneralisation of affixation rules. This kind of error should not always be considered as negative, because it suggests that L2 learners' cognition is continually active rather than language learners merely memorising as in rote learning.

5.1.4 The Intralingual and Interlingual Influences on the L2 Derivational Morphology

It might be quite difficult to recognise the sources of derivational morphological errors in essays that have been naturally produced by L2 learners, because one may need to specify whether those sources are derived from the L1 or L2, based on the contrastive analysis between the two languages. However, it seems to be generally acknowledged that intralingual influence is dominant whereas interlingual influence is limited, especially when the morphological system of learners' L1 (Korean) is quite different from that of L2 (English) (Jiang, 2000). Jiang (2000) noted that L2 learners are more likely to make inflectional errors in the L1 lemma mediation stage (the second stage) of his L2 vocabulary acquisition model (see Section 3.3.3) and suggested that there is no effect of L1 transfer in the morphological domain in the L2 mental lexicon. That is, according to Jiang (2000), there is only intralingual influence in the morphological domain, which means that L2 learners make morphological errors when they incorrectly apply the morphological rules of English. As mentioned in the previous section, McCutchen & Stull's (2015) concept of morphological invention is a good

example that shows this process, namely the overgeneralisation of derivational rules in the learners' mental lexicon. Tyler & Nagy (1989: 658) also pointed out that overgeneralisation is "the clearest type of evidence that acquisition of a morphological process as a combinatory, rule-governed process is taking place". This could also be supported by the fact that not only L2 learners but also English native speakers create words that do not exist in English. These kinds of morphological variations can sometimes become established as new words recognised by English dictionaries, e.g. *re-tweet*, *mini-stroke*, *hyper-local*, *de-friend* (O'Dell, 2015).

The derivational morphological errors made by L2 learners in previous studies could be interpreted in a very similar way to the 'morphological inventions' described by McCutchen & Stull (2015), in that they seem to be closely related to the overgeneralisation of affixation rules. In other words, the cognitive process involved in the production of derivational morphological errors in the L2 learners' mental lexicon is very closely associated with the activities of cognitive vehicles in the L2 lexical development model (see Section 3.4.2). Overall, as the current study seems to show, it appears that intralingual influence is a dominant source of the errors made in the L2 acquisition of derivational morphology.

On the other hand, interlingual influence is also a possible source of morphological errors, as when L2 learners seek to select appropriate affixes, their morphological knowledge can be associated with a literal translation process from L1. Studies that have investigated the interlingual influence (L1 transfer) in the L2 acquisition of morphology include Choi (2015), Lowie (2000), Cho & Tong (2014) and Lam & Sheng (2016). However, the studies seem to lack robust evidence of the connection between L2 morphological awareness and L2 learners' native language (L1). For example, Choi (2015) studied L1 Korean learners of English and claimed that L1 morphological awareness indirectly contributed to L2 reading comprehension.

However, she did not conduct a thorough comparison between the L1 (Korean) and L2 (English). Rather, she seemed to present superficial differences in the morphological systems of the two languages from a contrastive linguistic perspective and claimed that based on the learners' performance, the two variables correlated with each other. Choi did not specify which L1 morphemes attached to Korean words were associated with which English words, and she examined learners' morphological awareness of the L1 only, not the L2. In order to avoid these weaknesses, the current study uses a Korean reference corpus (see Section 6.2.4 for more information about the Sejong Corpus) as a norm to identify the interlingual influence shown in literal translations produced by L2 learners.

In summary, this study examines both intralingual and interlingual influence on derivational morphological errors detected from L2 learners' naturally occurring data (see Section 6.2.1). It first investigates the intralingual influence by detecting derivational morphological errors in learner corpus data through corpus-based analysis. It then speculates as to the interlingual influence by examining the degree of overgeneralisation of English affixation rules and comparing these errors with a Korean reference corpus (see Chapter 9).

5.2 Phonological/Orthographic Domain: Spelling Errors

In the phonological/orthographic domain, this study investigates spelling errors in a Korean learner corpus (see Section 6.2.1) in order to provide evidence of interlingual and intralingual influences in the L2 mental lexicon in the proposed new error taxonomy presented in Section 3.5.2. Just as Chapter 4 and the previous sections in this chapter have shown evidence of incongruity between L1 and L2 in the semantic, syntactic and morphological domains, this section reviews previous studies on spelling errors made by L2 learners and explores the

theoretical background for the interlingual and intralingual sources of spelling errors. It appears that interlingual spelling errors are affected by L1 phonology and/or orthography (Bebout, 1985; Ehri, 1993; Luelsdorff, 1986; Salam, 2016; Botley & Dillar, 2007; James, Scholfield, Carret & Griffiths, 1993), but it is difficult to determine whether the L1 orthographic influence is stronger or weaker than the L1 phonological influence, particularly for Korean and other learners of English whose L1 uses a non-Roman orthographic system (see Section 5.2.2). It is for this reason that the current study focuses on the L1 phonological influence rather than the L1 orthographic influence (see Section 5.2.3). On the other hand, intralingual spelling errors can also be detected and seem to be caused by the overgeneralisation of spelling rules (see Chapter 10).

5.2.1 Spelling Errors Made by L2 Learners

The goal of second language teaching and learning is to enable learners to communicate effectively in a given context. Knowing a language is not merely knowing the grammatical rules but also knowing how to express what one intends to express. Consequently, accuracy and fluency are two major elements that determine the success of language acquisition. In the current ESL/EFL environment, where the dominant approach is that of communicative language teaching (CLT), fluency tends to be prioritised whereas accuracy could be neglected. As James et al. (1993: 287) pointed out: “Prioritising communicativity has resulted in increased tolerance toward imperfection in second language users’ grammar and pronunciation.” Westwood (2014: 13) also claimed that even native English children make a lot of spelling errors due to the situation that still exists in literacy education where “[primary and secondary] students were expected to learn to spell merely by engaging in daily writing

and being exposed to print”. Similarly, L2 learners’ spelling errors have been neglected in recent years (Figueredo, 2006), even though orthographic representation is one of the major features that validate learners’ language proficiency.

In the 1980s and 1990s, however, a number of studies were conducted on spelling errors made by L2 learners. Many of them focused on the superficial description of orthographic structure to classify the spelling errors in the way L2 learners ‘insert’, ‘omit’, ‘substitute’ and ‘transpose’ incorrect letters of the given words (Pollock & Zamora, 1983), probably because the researchers were more interested in how the spelling errors could be differentiated on the surface rather than in the causes behind them. Recently, researchers have started to look at the importance of the factors that cause spelling errors. These factors have been classified into two types: L1 (interlingual) and L2 (intralingual) influence, which will be discussed in more detail in the next section.

5.2.2 Sources of Spelling Errors Produced by L2 Learners

As Adams (1994: 333) indicated, language users need to have “print awareness”; that is, they need to know the first steps to write or ‘print’ letters in words. Since there are several prerequisite elements to print letters appropriately (e.g. phonological awareness, letter and word recognition, phonemic awareness, conscious knowledge of sound-to-spelling correspondences), it might be very difficult to specify the sources of misspellings (Adams, 1994). Native speakers normally establish patterns of spellings and they may go through many developmental stages, sometimes making spelling errors in the process of the pattern formation (Al-Busaidi & Al-Saqqaf, 2015). If native speakers are liable to make spelling errors, L2 learners would be much more likely to do so. Furthermore, in ESL/EFL settings,

the mother tongue of L2 learners could also play a role in producing spelling errors. Since L2 learners have already acquired their L1 with its phonological and morphological systems, their L1 would inevitably affect the L2 acquisition of orthography. Accordingly, due to the interlingual influence, it seems more difficult to specify the sources of spelling errors produced by L2 learners. Indeed, researchers have generally agreed that there are two main sources of L2 learners' spelling errors, i.e. L1 (interlingual) and L2 (intralingual) sources. For example, Figueredo (2006) reviewed 27 previous studies on L2 learners' spelling skills and concluded that as well as providing evidence of interlingual influence on spelling errors, the studies also reported results where L2 learners tended to rely on English spelling rules, which might be overgeneralised to produce spelling errors (intralingual influence).

Fashola, Drum, Mayer & Kang (1996) identified these sources of spelling errors more specifically. They compared English spelling errors from dictation tasks produced by 38 Spanish children in the United States with the spelling errors produced by 34 native English speaking children. They divided the spelling errors of the Spanish elementary students into two categories: 'nonpredicted errors' and 'predicted errors'. The nonpredicted errors were random spelling errors that might be caused by a lack of knowledge of English spelling rules. For example, the L2 learners were likely to overgeneralise 'English spelling rules' in the process of inflection or derivation. The predicted errors were produced when the Spanish learners of English applied 'Spanish spelling rules' that were inappropriate for English, an interference from the L1 that Fashola et al. (1996) considered predictable and to be expected. Note that the predicted spelling errors could be caused by two separate factors: L1 phonology and L1 orthography. With regard to phonology, some sounds exist in English but not in Spanish (e.g. [sh] in *shoe*), where Spanish learners of English might have substituted alternative Spanish sounds for English sounds. In the case of the orthographic interference,

Spanish learners might apply Spanish orthographic rules to English spellings, because although English and Spanish have different degrees of sound-to-spelling correspondences, the two languages use an almost identical alphabetic system.

It could be argued that the unpredicted and predicted spelling errors from Fashola et al.'s (1996) study can be regarded as identical to the intralingual and interlingual errors respectively in the current study, in that the spelling errors from each category in both studies are derived from the same language sources (L2 and L1 respectively). As explained above, the intralingual spelling errors are produced as the result of L2 learners' overgeneralisation of English spelling rules and consequently these errors can be found in the writings of L2 learners with various L1 backgrounds. On the other hand, the interlingual spelling errors depend on the specific L1 background and can therefore be expected to be diverse. These errors are also more difficult to identify, because the error detection should take into account not only a contrastive analysis between the L1 and L2 but also the phonological and orthographic processing skills of L2 learners.

Other studies that have examined L1 phonological interference as a causative factor in L2 learners' spelling errors include Bebout (1985), Ehri (1993), Luelsdorff (1986), Salam (2016), Arab-Moghaddam & Sénéchal (2001), Botley & Dillar (2007), El-Hibir & Al-Taha (1992), Panah & Padakannaya (2008) and James et al. (1993). Salam (2016) and El-Hibir & Al-Taha (1992) reported that Arabic learners of English tend to have difficulty with English spellings with 'p' and 'b', because whereas in English there are two bilabial stops ([p] and [b] sounds), in Arabic there is only one ([b] sound), e.g. **batch* (<*patch*>), **pobular* (<*popular*>). In a study with L1 Welsh learners of English, James et al. (1993) showed that the L1 influence could account for up to 38.5% of the spelling errors in their data, which suggested that the interlingual influence on misspellings could be very powerful in certain cases.

There have also been studies of L1 orthographic interference (e.g. Cook, 1997; Figueredo, 2006; Fashola et al., 1996; James et al., 1993). The spelling errors caused by L1 orthography have been considered as related to the Orthographic Depth Hypothesis (ODH) as well as to orthographic differences between the L1 and L2. According to Frost (1994), orthographic depth refers to the degree of transparency of the sound-to-spelling correspondence and this can be described as ‘deep’ or ‘shallow’. Since different languages have different writing systems, the transparency of the relation between spelling and phonology varies from one language to another. For example, languages such as Spanish, Italian, Serbo-Croatian and Korean have shallow orthography, whereas English has a deep orthography in which there is a more complex or opaque sound-to-spelling relation (Lee, 2010; Hamada & Koda, 2008; Wang & Geva, 2003). The ODH suggests that learners whose first languages are shallower, like Spanish or Italian, may have greater difficulty in learning English’s opaque orthography (Figueredo, 2006), since they all use very similar Roman orthographic systems. Thus, the spelling errors produced by Spanish or Italian learners of English could be caused by the similar but slightly different Roman orthographic systems. Meanwhile, L2 learners of English whose L1s do not use Roman orthographic systems, such as Chinese, Japanese and Korean, might experience different cognitive processing when they decode and encode English scripts, which may cause more or fewer spelling errors (Barnitz, 1982). Wang & Geva (2003: 2) claimed: “Recent research suggests that the orthographic depth framework can be extended to nonalphabetic writing systems such as Chinese” and this might also apply to the Korean language.

It is also worth noting that some languages that do not use the Roman alphabetic system as their primary writing system do have a Roman alphabetic equivalent system (the use of Roman script for L1 languages), e.g. ‘pinyin’ for Chinese and ‘romaji’ for Japanese. In these

cases, the equivalent systems could also play a role in spelling errors. For example, Cook (1997) reported that Japanese learners of English made spelling errors such as *sarary* for *salary*, *grobal* for *global*, and speculated that these errors were probably due to interference from the romaji script, in which each Japanese letter is mapped with the corresponding letter from the Roman alphabet. In the case of loan words from English, such as *salary* and *global*, the conventional romaji spellings might be different from the original English spellings and this could cause Japanese learners to make spelling errors.

However, the spelling errors produced by Korean learners of English in the current study seem to be quite different from those of the Japanese learners in Cook's (1997) study, not only because Korean has a more transparent or shallower orthography than English, but also because the Korean orthographic system, *hangeul*, is totally different from the Roman alphabetic writing system. Moreover, Korean does not use a Roman alphabetic equivalent system such as pinyin or romaji. Therefore, the current study focuses more on Korean phonological influence that could cause Korean learners of English to make interlingual spelling errors, rather than on Korean orthographic influence.

5.2.3 The Interlingual and Intralingual Spelling Errors Made by Korean Learners of English

This study hypothesises that the incongruency between L1 and L2 can also be seen in the phonological/orthographic domain. Through a corpus-based error analysis of learner corpus data produced by Korean learners of English (see Section 2.2.1), it shows spelling errors caused by both interlingual and intralingual influences. These spelling errors provide

evidence of the interlingual and intralingual influences in the phonological/orthographic domain in the L2 mental lexicon in the proposed new error taxonomy (see Section 3.5.2).

A number of studies have reported spelling errors made by Korean learners of English. The majority of them focused only on the superficial orthographic structure of misspelled words and classified the errors into categories such as *insertion*, *omission*, *substitution* and *transposition* (e.g. Jeong, 2013; M. Park, 2015; Kim & Son, 2012; Moon & Kim, 2015). However, these studies did not address the sources or causes of the spelling errors. To the best of my knowledge, to date no study has investigated the sources of spelling errors made by Korean learners of English. One reason for this might be that it would be difficult to determine whether the mismatch of orthographic and/or phonological correspondences between Korean and English could play a role in spelling errors, because the Korean language uses its unique orthographic system *hangeul*, as mentioned in Section 5.2.2. There have been contrastive analyses between Korean and English phonemes, but not between the two systems of orthography (e.g. Heo, 2004; Jeong & Cho, 2016). Some studies, like Lee (2004) and Lee (2013) have reported that Korean learners of English have difficulty in pronouncing specific English vowel sounds appropriately. In particular, Lee (2007) analysed spelling errors made by Korean elementary students and some of the errors reported in that study seemed to suggest that the students sometimes failed to connect certain L1 (Korean) phonemes to appropriate L2 (English) phonemes, which might result in L2 spelling errors. However, it would be very difficult to specify the sources of the interlingual spelling errors with respect to the correspondence of vowel sounds and letters between Korean and English, due to the complexity and richness of the vowel system and its many variations (Bebout, 1985; Moon & Kim, 2015). That is to say, one vowel letter can be pronounced in many ways depending on the letters it is combined with, and conversely, one vowel sound can be

represented orthographically in English by many different combinations of letters. One might be able to investigate spelling errors related to vowels through experimental studies, in which one could control the variables of specific vowel and letter combinations. However, that does not seem appropriate for a corpus-based study that deals with naturally occurring data. Therefore, this study examines consonant spelling errors, in which the sound-to-spelling correspondences are much clearer than with vowels.

Table 5.2 shows the correspondences between English and Korean consonant phonemes, based on the place and the manner of articulation (Lee, 1997). Although almost every English consonant has corresponding Korean phonemes and letters, there are several phonemes that do not have equivalent sounds and letters, as indicated by the shaded boxes. For Korean learners, the phonemes [š] and [ž] may be unfamiliar sounds but are unlikely to be problematic in spellings that represent them because there are no equivalents or similarly recognised consonants in Korean. Korean learners are likely to map them to the letters of alveolar sounds, e.g. [ㅅ], [ㅆ]. Likewise, it seems that Korean learners may have difficulty in recognising the two dental sounds [θ] and [ð], but not with writing spellings that represent these sounds. Because there are no equivalent sounds in Korean and no letters that confuse Korean learners, there could be no interference. However, the three phonemes [f], [v] and [r] and their corresponding letters can be much more problematic for Korean learners of English, in terms of the sound-to-spelling correspondences. This is why the current study focuses on these three specific kinds of consonant spelling errors in order to show the interlingual influence of L1 phonological mediation on L2 orthography. In other words, just as with the other sounds (e.g. [š], [ž], [θ] and [ð]), the Korean language has neither sounds nor letters that are equivalent to the English [f], [r] and [v]. However, unlike the other sounds, it seems that

Korean learners tend to confuse these three sounds with Korean [p], [l] and [b] sounds respectively, and therefore substitute them (see La, 2001; Lee, 2007; Jeong & Cho, 2016; Ha, Johnson & Kuehn, 2009). Consequently, this study hypothesises that Korean learners of English might make the corresponding spelling errors, i.e. substituting ‘f’ or ‘ph’ for ‘p’, ‘r’ for ‘l’, ‘b’ for ‘v’ and vice versa.

Table 5.2 Comparison between English and Korean consonant systems (adapted from Lee, 1997: 30)

Place of Articulation Manner of Articulation		Bilabial	Labiodental	Dental	Alveolar	Palato-alveolar	Velar	Glottal
Stops	English	p b			t d		k g	
	Korean	ㅂ ㅂㅂ ㅍ			ㄷ ㄷㄷ ㅌ		ㄱ ㄱㄱ ㅋ	
Fricatives	English		f v	θ ð	s z	š ž		h
	Korean		-	-	ㅅ ㅆ ㅆ	-		哨
Affricates	English					č ĥ		
	Korean					ㅈ ㅉ ㅊ		
Nasals	English	m			N		ŋ	
	Korean	ㅁ			ㄴ		ㅇ	
Liquids	English			L	r			
	Korean			ㄹ	-			
Glides	English	w			y			
	Korean	₩						

One important concern that should be mentioned here regards the allophones or variation of sounds of the three consonant letters. Each of the letters can be pronounced slightly differently depending on their position as initial, middle or final letters, or on the adjacent letters, although the sounds are phonetically categorised as identical. For example, the [l] phoneme has a clear sound when it precedes a vowel, as in *listen* or *fall in*. On the other hand,

it has a somewhat devoiced sound when preceded by a voiceless plosive as in *please* or *clue* and a dark sound when it occurs as the final letter in a word after a vowel, as in *fall down* (Roach, 2000). However, the effect of allophonic variance is limited and so can be ignored in terms of the sound-to-spelling correspondences of two languages, because language learners might not be able to perceive the difference of phonetic sounds (Lee, 1997). Even if they could perceive the subtle differences between sounds, this would be so slight that they would not select other letters to represent those sounds. In order for language users (including learners) to perceive the phonetic sounds and to map them with their corresponding letters, their psychology of speech sounds is more important than the actual speech sounds in a linguistic context (Sapir, 1925). As Carr (1999: 38) put it, these similar sounds are “phonetically distinct” but “phonologically equivalent”. In other words, even native English speakers would perceive the slightly different [l] sounds of *listen*, *please* or *fall down* as identical, because they have a psychologically determined concept of the phoneme [l] in the English phonological patterns (Chomsky & Halle, 1991).

On the other hand, as mentioned in Section 5.2.2, if Korean learners also overgeneralise the spelling rules of English, this might result in intralingual spelling errors. In this regard, the current study investigates only those misspellings that are made within the process of inflection and derivation. This is because when L2 learners use inflectional and derivational morphemes, one can see whether they apply spelling rules correctly or incorrectly. For example, they could make spelling errors while adding inflectional morphemes (e.g. present progressive *-ing*, plural *-(e)s*, regular third person singular *-s* and regular past tense *-(e)d*), or derivational morphemes (e.g. affixes like *dis-*, *mis-*, *un-*, *-ly*, *-able*). Note that the derivational spelling errors do not include incorrect morphological selections, which belong to the morphological domain (see Section 5.1.3). For example, errors such as

**discommunication* (<*miscommunication*) or **harmness* (<*harm*) are not categorised as intralingual spelling errors because they are more likely to derive from the incorrect selection of specific affixes (*dis-*, *-ness*), rather than from the overgeneralisation of English spelling rules. Conversely, errors like **useing* (<*using*), **teached* (<*taught*), **easilly* (<*easily*) and **absolutly* (<*absolutely*) are classified as intralingual spelling errors, because they seem to result from the overgeneralisation or incorrect application of spelling rules within the process of inflection or derivation, i.e. insertion and omission of unnecessary letters, or incorrect application of inflectional rules. This category also includes words that contain incorrect spellings of attached inflectional or derivational morphemes.

In summary, this thesis shows the empirical evidence detected through a corpus-based analysis and discusses how the results in the phonological/orthographic domain relate to the proposed new error taxonomy presented in Section 3.5.2.

5.3 Chapter Summary

This chapter has aimed to provide a theoretical basis for the error types from the morphological and phonological/orthographic lexical domains, according to the new error taxonomy proposed in this study.

In Section 5.1, I explained the concept of morphological awareness and drew particular attention to the distributional aspect of derivational morphology in order to show that the overgeneralisation of affixation rules plays a role as intralingual influence in the L2 acquisition of morphology. I also discussed the possible interlingual influence in the morphological domain in the process of literal translation from Korean to English.

In Section 5.2, I reviewed previous studies that have reported both interlingual and intralingual spelling errors. In terms of interlingual influence, Korean learners could be affected by the mismatch between Korean and English phonology regarding certain consonants. Intralingual influence, on the other hand, could cause inflectional or derivational spelling errors where Korean learners overgeneralise or incorrectly apply the English orthographic rules.

Given the theoretical background to the error features from the four domains of the new error taxonomy presented in Chapters 4 and 5, the next chapter will address the methodological basis of the study, including research questions, data (a learner corpus and reference corpora), research techniques and procedures of the corpus-based error analysis.

CHAPTER 6: METHODOLOGY

This chapter sets out the methodological basis for the research presented in this thesis. As mentioned, the study is a learner corpus research in which researchers need learner corpora as learner data and a corpus software program that is utilised for data analysis. As Granger (2002: 12) noted, “[LCR] focuses on errors in interlanguage and uses computer tools to tag, retrieve and analyse them”. Especially in the current study, reference corpora are used as a norm of standard English, and a concordancing package (*WordSmith Tools*) makes it possible to efficiently process and analyse a large amount of learner data in order to detect the deviant expressions or errors in the four lexical domains.

In Section 6.1, I state the two main research questions and sub-questions of this study and indicate how these questions are to be answered in more detail. The research questions are related to the four main results of the study, which support the grounds for the L2 lexical development model and the new error taxonomy I proposed in Sections 3.5.1 and 3.5.2. In Section 6.2, I introduce the corpora used in the study. These comprise a learner corpus consisting of essays written by Korean learners of English, English reference corpora and a Korean reference corpus, which are used to detect learners’ errors and to identify interlingual or intralingual influence that might cause these errors. In Section 6.3, I explain the preliminary examination for selecting the four main error features and outline the different research procedures and specific techniques to detect each error feature in the four lexical domains.

6.1 Research Questions

As already stated in Section 1.3, this study aims to answer two main research questions (RQs):

RQ1. How do interlingual and intralingual influences affect the production of L2 learners' lexical errors?

RQ2. Is it possible to categorise the lexical errors according to their sources and domains presented in the proposed new error taxonomy?

In order to address the research questions, first, I have already proposed the L2 lexical development model as a conceptual framework that includes four lexical domains in the L2 mental lexicon (see Section 3.5.1). I have also proposed a new error taxonomy based on this model. RQ1 is related to the main results of this research, which are discussed from Chapter 7 to Chapter 10. Corpus-based methods were utilised to investigate whether and how the lexical errors are produced in each domain. More specifically, I selected one feature from each domain to identify the interlingual and intralingual influences on the lexical errors. RQ2 can be considered as a follow-up question, because it is closely related to whether or not the L2 lexical development model and the new error taxonomy are well-grounded: the question which can be verified by answering RQ1. RQ1 (and RQ2) can be answered by addressing the following four sub-questions, one for each feature:

RQ1-1. How do the interlingual and intralingual influences affect the production of collocational errors of dimensional adjectives?

RQ1-2. How do the interlingual and intralingual influences affect the production of over-passivisation errors of non-alternating unaccusative verbs?

RQ1-3. How do the interlingual and intralingual influences affect the production of derivational morphological errors?

RQ1-4. How do the interlingual and intralingual influences affect the production of spelling errors?

In order to address RQ1-1, I used three pairs of dimensional adjectives: *large-small*, *high-low* and *long-short*. All the concordance lines in the YELC (see Section 6.2.1) that contain the six dimensional adjectives, including comparative forms as well as base forms, were extracted. The Bank of English (BoE) and the British National Corpus (BNC) were utilised as reference corpora (see Sections 6.2.2 and 6.2.3) as a norm to identify whether or not the collocations of dimensional adjectives with the modified nouns should be judged as errors. In the next chapter (Chapter 7), the underlying possible sources of these dimensional adjective errors will be discussed from a cognitive linguistic perspective.

In order to address RQ1-2, first, I chose six English non-alternating unaccusative verbs: three ‘matched’ and three ‘mismatched’ verbs according to their Korean equivalents (see Sections 6.3.2). If L1 morphology as the interlingual influence plays a role in the L2 acquisition of unaccusative verbs, then the error frequency of mismatched non-alternating verbs will be greater than that of matched verbs. Because the Korean equivalents of mismatched English non-alternating unaccusative verbs have their causative counterparts, we would expect L2 learners to presume that the mismatched English verbs can be passivised or can carry objects just as the Korean equivalents can (see Section 4.2). In addition, I compared the error frequency of the six English non-alternating unaccusative verbs in each sub-corpus according to learners’ proficiency levels.

Secondly, I investigated error frequencies depending on whether subjects in the sentences are animate or inanimate in order to identify the intralingual influence. To clarify whether or not Korean learners make over-passivisation errors of non-alternating unaccusative verbs when subjects are animate or inanimate, I compared the numbers of correct and incorrect use of the verbs in a statistical manner (see Table 6.3 in Section 6.3), according to the subject animacy (see Chapter 8).

To answer RQ1-3, I searched for derivational morphological errors in the YELC and divided these into four categories according to the degree of deviance in the association of root morphemes and affixes. Derivational morphological errors in the concordance lines were closely examined to check whether some might be literal translations from Korean. I referred to a Korean standard dictionary and a Korean reference corpus, the Sejong Corpus (see Section 6.2.4) in order to identify the interlingual influence. In addition, I investigated whether Korean learners have more difficulty with certain types of derivational affixes. To do so, I divided the affixes into two different kinds, class-maintaining or class-changing affixes, and counted the derivational errors of each. Finally, I compared the error frequency of each proficiency group from the YELC, in an attempt to show how the distributional knowledge of derivational morphology develops as the learners' proficiency improves from low to intermediate, and to advanced level (see Chapter 9).

In order to answer RQ1-4, I focused particularly on specific spelling errors that clearly show the interlingual and intralingual influences, i.e. the spelling errors of three consonant letters for the interlingual influence and the overgeneralisation of spelling rules in the process of inflection and derivation for the intralingual influence. In addition, I compared the error frequencies of each proficiency group from the YELC. This helped to show how the

orthographic knowledge develops as the learners' proficiency improves from low to intermediate, and to advanced level (see Chapter 10).

6.2 Corpora

This research adopted a corpus-based error analysis, using five corpora. In this section I briefly introduce these corpora with regard to their characteristics in terms of size, data type and method of use:

- 1) The Yonsei English Learner Corpus (YELC)
- 2) The Bank of English (BoE) as the main English reference corpus
- 3) The British National Corpus (BNC) and *MorphoQuantics* as supplementary English reference corpora
- 4) The Sejong Corpus as a Korean reference corpus

The YELC contains the learner data that were analysed in this study. The BoE is the main English reference corpus for this research, used as an English norm to determine learners' errors in the YELC. Another reference corpus, the BNC, was used to triangulate the learners' data in the YELC, especially when the BoE did not clearly show deviance between the errors and the English norm. *MorphoQuantics* contains complex word types extracted from the BNC (Laws & Ryder, 2014) and was referred to when selecting relevant affixes to be investigated in the analysis of derivational morphological errors (see Section 6.3.3). Then, a Korean reference corpus, the Sejong Corpus, was used as a Korean norm to identify the interlingual influence. The following sections describe these corpora in more detail.

6.2.1 The Yonsei English Learner Corpus (YELC)

The YELC was compiled in 2011 by Yonsei University, Seoul. Founded in 1885 by an American missionary, Yonsei is now considered one of South Korea's most prestigious universities (Rhee & Jung, 2012). In the 2009 Academic Ranking of World Universities the university was placed third in South Korea and in 2012 it was situated in the 23-42 range in the Asia/Pacific region. The student population includes approximately 25,000 undergraduates and 12,000 postgraduates. The YELC, established thanks to funding from the government project Brain Korea 21, comprises narrative and argumentative essays of first year undergraduates (mainly 19-20 years old) of the university, all of whom had recently graduated from high schools and whose English skills vary from beginner or intermediate level to advanced level. The YELC was chosen for this study because:

- 1) It is easily accessible. The whole data of the YELC can be gained with permission. It is also one of very few learner corpora available that comprise materials by Korean learners of English. Owing to the lack of available learner corpora, researchers do not often adopt learner corpus research (LCR), despite an increasing interest in naturally occurring data.
- 2) The YELC is very large (approximately 1 million words), which means that the target words for the current study occur frequently. It is more extensive than the ICLE (International Corpus of Learner English), which may be considered as a prototypical learner corpus because it is a pioneering work that contains highly homogeneous L2 learners' writings from several mother tongue backgrounds (Granger, 2003): while the sub-corpora of the ICLE comprise 100,000 to 200,000 words written by learners of various L1 backgrounds, the YELC has about 1 million words produced by Korean learners (for more information about the YELC, see Appendix 1).
- 3) One of the merits of the YELC is that every essay is graded into one of nine proficiency levels according to the refined version of the CEFR (the Common European Framework of Reference for Languages) writing scales. This makes it possible to examine the error rates of specific error features according to proficiency

level. Table 6.1 shows the details of the YELC (for data samples from each level, see Appendix 2).

Table 6.1 Numbers of words and texts in the YELC according to proficiency level

Level	Grade	Number of texts	Number of words
Low	A1	82	3,056
	A1+	370	36,009
	A2	1,368	195,473
	Sum	1,820	234,538
Intermediate	B1	2,346	391,463
	B1+	1,410	263,470
	B2	756	146,843
	Sum	4,512	801,776
Advanced	B2+	162	33,250
	C1	74	15,434
	C2	4	829
	Sum	240	49,513
Total		6,572	1,085,827

In addition, it should be noted that the data in the YELC are naturally occurring data. This gives researchers a different perspective to interpret the results of studies in L2 acquisition. It has been argued that, in general, in the process of L2 acquisition, productive skills (speaking and writing) are more difficult to acquire than receptive skills (listening and reading) (Nation, 2001; Ellis & Beaton, 1993). Explanations for why productive skills involve a greater learning burden take into account aspects such as amount of knowledge, practice, access and motivation (Nation, 2001). Whatever the reasons might be, it seems clear that the acquisition of productive skills requires more time and effort from L2 learners than does the acquisition of receptive skills.

With regard to the four features considered in the current research, it is very hard to find studies that have dealt with productive data. For example, although a number of studies have looked at the L2 acquisition of unaccusative verbs (see Section 4.2), most have dealt with the subjects' receptive skills, especially reading. The data in these studies were usually collected through grammaticality judgement tests (or picture judgement tests in a few cases); however, some researchers have questioned the reliability of the data elicited by these kinds of tests. According to Han (2006), one of the major weaknesses of grammaticality judgement tests is subjects' indeterminacy; that is, since the subjects are forced to choose either yes or no, it is probable that they will guess the answers even when they are not sure. In that case, they might be able to use performance strategies rather than language competence; for example, having a balance between the number of grammatical and ungrammatical answers, or avoiding a certain number of consecutive answers. As Ellis (1991: 164) indicated, “[...] grammaticality judgment may reflect processing factors. If this is so, it is dangerous to make claims about the nature of the learners' competence on the evidence provided by grammaticality judgments.” Therefore, grammaticality judgement tests should be used with caution. In contrast, naturally occurring data preclude not only weaknesses derived from grammaticality test items but also the likelihood of guessing or the use of performance strategies by subjects.

Indeed, some studies have tried to collect productive data. Zobl (1989) collected only a small data set, mainly from Japanese college students, where the token¹⁴ of the verbs is only 246 occurrences. Balcom (1997) conducted a controlled productive task, a cloze test that asked subjects to decide between a passive and an active form, given a base form of the verb. However, the data elicited by such methods are still far from learners' spontaneous writings.

¹⁴ In corpus linguistics, the term 'token' refers to the total number of words, whereas the term 'type' refers to the number of different words (Hunston, 2002).

A very small number of studies have investigated learner corpus data (e.g. Oshita, 2000; Shin, 2011), which are authentic writings of L2 learners of English. However, even these studies only report the distribution of different error types of unaccusatives and do not account for their possible causes.

Similarly, as mentioned in Section 5.1.1, very few studies of morphological awareness have employed naturally occurring data. Instead, the majority of the previous studies have been carried out using reading comprehension or written elicitation tasks. Moreover, although distributional knowledge is the most challenging among the three components of morphological awareness, previous studies have not dealt with this aspect (see Section 5.1.2). Consequently, there is a need to analyse productive learner data. The reason for the relative lack of studies dealing with productive data may be the difficulty of collecting naturally occurring data. In employing such data, the current study offers a broader perspective to account for derivational morphological errors in relation to distributional knowledge of derivational morphology.

To analyse the data in the YELC, this study utilises *WordSmith Tools 5.0* (Scott, 2012). This well-known and widely used software provides three main tools: Concord, Keywords and Wordlist. Concord makes concordance lines of a specified search word and gives access to information about collocates, dispersion plots, etc. Keywords seeks out the ‘key’ words whose frequencies are, statistically, unusually high (keyness) in comparison with some norms. Wordlist generates lists of words from one or more text files, by ranking according to frequency or other statistical criteria (e.g. *t*-score, log-likelihood). In the analyses of the YELC in the current study, Concord is mainly used to retrieve dimensional adjectives, non-alternating unaccusative verbs and derivative words (see Sections 6.3.1, 6.3.2 and 6.3.3), while Wordlist is used to identify spelling errors (see Section 6.3.4).

6.2.2 The Bank of English (BoE): the Main English Reference Corpus

As already stated, this research used the Bank of English (BoE) as the main English reference corpus in order to investigate the collocational behaviour of dimensional adjectives in the semantic domain. The BoE,¹⁵ based at the University of Birmingham, is one of the largest general English corpora, with the latest version, released in 2002, containing 450 million words. The enormous size of the BoE is the most important reason for using it as the main reference corpus in this study: size is a crucial consideration, particularly when investigating the collocational behaviour of sequences of words (Sinclair, 1991, 2004), because it would be very difficult to obtain reliable frequency figures for given collocations from a small corpus. Walker (2008) showed how the frequency of a particular sequence with the word *aspect* dropped by one digit with an additional word. In the BoE, the single word *aspect* occurs 12,482 times and the two-word sequence *aspect of* occurs 9,007 times, whereas for the three-word sequence *one aspect of*, the frequency drops to 796. Although this example is something of a ‘rule of thumb’, it clearly shows the need for a very large corpus for a collocation study. The easy access to the BoE available to the researcher as a postgraduate student of the University of Birmingham is another advantage of using this corpus.

Table 6.2 shows what kinds of texts are included in the BoE. The sub-corpora of the BoE are made up of around 70 percent British English (30 percent American English or other) and around 90 percent written English (10 percent spoken English), which might be seen as unbalanced in terms of register. The BoE is also sometimes criticised because it is dominated by journalistic sources (Walker, 2008). However, these considerations seem not to present a problem for the current study, given the enormous size of the corpus.

¹⁵ More information about the Bank of English is available at: <http://www.titania.bham.ac.uk>.

Table 6.2 The sub-corpora of the Bank of English (from Barnbrook, Mason & Krishnamurthy, 2013: 214)

Sub-corpora name	Size (the number of token)	Country of origin	Contents
bbc	18,604,882	UK	BBC radio
brbooks	43,367,592	UK	general books
brehem	4,640,529	UK	ephemera
brmags	44,150,323	UK	magazines
brspok	20,078,901	UK	informal spoken language
econ	15,716,140	UK	<i>Economist</i> magazine
guard	32,274,484	UK	<i>Guardian</i> newspaper
indy	28,075,280	UK	<i>Independent</i> newspaper
newsci	7,894,959	UK	<i>New Scientist</i> magazine
npr	22,232,422	US	public radio
oznews	34,940,271	Australia	newspapers
strathy	15,920,137	Canada	Canadian mixed corpus
sunnow	44,756,902	UK	<i>Sun/News of the world</i> newspaper
times	51,884,209	UK	<i>Times/Sunday Times</i> newspaper
usacad	6,341,888	US	academic books
usbooks	32,437,160	US	general books
usephem	3,506,272	US	ephemera
usnews	10,002,620	US	newspapers
usspok	2,023,482	US	informal spoken language
wbe	9,648,371	UK	business language

In order to gain access to the BoE, the established Telnet protocol is utilised to log into the corpus server and use the text-based search and analysis software (*LookUp*) that is part of the BoE package. The interface of the *LookUp* program might look quite old-fashioned, and is currently being updated. However, the program provides not only basic functions such as retrieving concordance lines with KWIC (Key Word In Context) presentation, but also the

picture function (or picture profile) that shows an overview of the collocates within a span of three to six, ranking by frequency, mutual information or *t*-score. The picture function of *LookUp* is a very powerful command to show the relationship between a node and its collocates, which is very useful for searching collocations (see Figure 7.2 in Section 7.2 for an example of the picture function).

6.2.3 The British National Corpus (BNC) and *MorphoQuantics*: Supplementary English Reference Corpora

Another reference corpus used in this study is the BNC, which contains 100 million words of texts from a wide range of genres. The BNC was originally created from 1991 to 1994 and is regarded as a ‘balanced’ general English corpus in terms of the variety of its contents, whereas the BoE is journalistic. Similar to the BoE, 90 percent of the BNC content is written text. However, this includes not only newspapers and journals, but also popular fiction, letters and memoranda, among others. The spoken section that makes up the remaining 10 percent includes 863 transcripts of, for example, informal conversations, radio shows and business meetings (McEnery et al., 2006). One of the reasons why the BNC is widely used as a reference corpus is the powerful corpus-analysis interface of the *BNCweb*, which enables users to retrieve the collocates and display distributional information, of the lexical item.

*MorphoQuantics*¹⁶ is a specialised corpus of complex words from the spoken part of the BNC. It contains a comprehensive set of 17,943 complex word types classified by prefixes, suffixes and combining forms as proposed by Stein (2007). According to Laws & Ryder (2014), *MorphoQuantics* was developed with the aim of providing the type and token frequencies of

¹⁶ It can be accessed at: <http://morphoquantics.co.uk>.

the components of complex words according to affixes. It includes 835 affixes in total and these are also categorised from an etymological perspective. The affixes investigated in the current study were chosen with reference to this kind of information provided by *MorphoQuantics* (see Section 6.3.3).

6.2.4 The Sejong Corpus: the Korean Reference Corpus

In order to identify the interlingual influence on collocational errors of dimensional adjectives and derivational morphological errors, this study used a Korean reference corpus. The Sejong Corpus was developed by the Korean government from 1997 to 2011 (Hong, 2009; Hwang & Choi, 2016) and has a written section that consists of approximately 37 million words. In order to retrieve specific words in the corpus, I used a web-based corpus tool, the KKMA¹⁷ (Lee, Yeon, Hwang & Lee, 2010). The Korean language has an abundance of morphemes, which is one of the features of the agglutinative languages, and so it needs a specialised corpus tool for the analysis. The existence of Korean equivalents in the Sejong Corpus that can be literally translated into the errors in English is considered to be evidence of interlingual influence on collocational errors of dimensional adjectives and derivational morphological errors (see Sections 6.3.1 and 6.3.3).

6.3 Research Procedures

This section describes the research methods and procedures whereby corpus-based analysis is used to investigate the following four error features: collocational errors of dimensional adjectives (Section 6.3.1), over-passivisation errors of non-alternating unaccusative verbs

¹⁷ It can be accessed at: <http://kkma.snu.ac.kr.d>

(Section 6.3.2), derivational morphological errors (Section 6.3.3) and orthographic (spelling) errors (Section 6.3.4).

As already mentioned in Section 3.5.2, this study does not aim to establish an exhaustive and comprehensive error taxonomy that includes every kind and aspect of Korean learners' lexical errors in English. Rather, it intends to show one representative error type for each of the four domains, using corpus-based error analysis. The YELC does contain other error features that could be classified into the four domains but are not selected for investigation in this study, usually because the error frequency is too low to be relevant for a corpus-based analysis or for statistical measurement. Therefore, before embarking on the main work for this thesis, I conducted a preliminary examination in order to determine which features of errors needed to be investigated.

The data for the preliminary examination were randomly chosen writing samples from the YELC. These comprised 228 text files (32,528 words). First, I investigated the writing samples and manually extracted any possible types of lexical errors. Then, I consulted the reference corpora (the BNC and the BoE) to check whether or not the potential errors were deviant. If the errors seemed to be deviant and worth further investigation, they were divided into provisional categories that were refined later on. Once an error category was set up, more errors were collected from the whole data of the YELC through the automatic extraction used by *WordSmith Tools*. If a fruitful number of errors were collected in a category and they seemed to show possible causes from a cognitive linguistic perspective, then they became candidates for further investigation. Otherwise, the category was discarded from the current research, although the errors in that category might still be worth investigating by future research with much larger datasets. Through this circular process, and informed by literature

reviews, the error categorization became much clearer and finally the four error features were selected.

Table 6.3 Outline of the research of the four error features

	Search items	Corpora used in the analysis	Methods & procedures
Dimensional adjectives	- Three pairs of adjectives (<i>large-small, high-low, long-short</i>)	- The BoE - The BNC - The Sejong Corpus	- Errors detected using <i>WordSmith Tools</i> - Compared with reference corpora
Non-alternating unaccusative verbs	- Three matched verbs (<i>exist, happen, arrive</i>) - Three mismatched verbs (<i>occur, remain, appear</i>)	N/A	- Errors detected using <i>WordSmith Tools</i> - Statistical analysis (error rates, chi-square test)
Derivational morphology	- Eight class-changing affixes (<i>-able, -ible, -er, -less, -ish, -ly, -ness, -ize(ise)</i>) - Seven class-maintaining affixes (<i>un-, in-, im-, il-, ir-, dis-, re-</i>)	- <i>MorphoQuantics</i> - The Sejong Corpus or the Standard Korean Language Dictionary ¹⁸	- Errors detected using <i>WordSmith Tools</i> - Errors divided into four types - Statistical analysis (error rates, chi-square test, G square test) - Compared with the reference corpus
Orthography (Spelling)	- Three pairs of consonant sound substitution ([p]-[f], [l]-[r], [b]-[v]) - Inflectional and derivational spelling errors	N/A	- Errors detected using <i>WordSmith Tools</i> (Match function with Texts files of the BNC World corpus) - Statistical analysis (error rates, chi-square test)

Since the research methods and procedures differ slightly among the four error features, albeit that they all follow a corpus-based error analysis, Table 6.3 outlines the methodology of each piece of research, which will be described in more detail in the following sections.

¹⁸ It can be accessed at the URL: <http://stdweb2.korean.go.kr/main.jsp>.

6.3.1 Research Procedures: Collocational Errors of Dimensional Adjectives

It is important to note that the procedures to identify collocational errors of dimensional adjectives are closely related to those used to discover the degree of deviation of collocations, as mentioned in Section 4.1.4.

Three pairs of dimensional adjectives in the YELC were analysed: *large-small*, *high-low* and *long-short*. The main reason for this choice is that they occur frequently in the YELC. It might also be possible to examine errors of other dimensional adjectives, such as *wide-narrow*, *deep-shallow*, if one were to collect a larger dataset containing a considerable number of such dimensional adjectives. However, in the YELC, the frequencies of these dimensional adjectives were not high enough to yield fruitful results: the word *wide* occurs only 38 times, *narrow* 13 times, *deep* 79 times and *shallow* 4 times.

WordSmith Tools 5.0 was used to extract all the concordance lines of the three pairs (six) of dimensional adjectives. Then, the deviant associations were compared with data from reference corpora, i.e. the BoE and the BNC. In order to extract concordance lines of the collocations of dimensional adjectives, the search item ‘*adjective+0,1noun@*’ was entered at the query page in the *LookUp* program. For example, I entered ‘*large+0,1damage@*’ as a command in order to obtain the concordance lines of associations of the noun *damage* with the pre-modifying dimensional adjective *large*. By using the command, I was able to collect all the concordance lines that include the co-occurrences of *large* and the lemma *damage* and allow up to one ‘slot’ (a space for a word) to appear between the two items, such as ‘*large damage*’, ‘*large damages*’, ‘*large ... damage*’. By increasing the number of slots between the adjective and the noun, it is possible to obtain more concordance lines of given associations. However, if the number of slots is two or above, then one may obtain too many concordance

lines, including ones where the adjective is not pre-modifying the noun, and these should be discarded manually. Moreover, when using two or more slots, I rarely (or never) found relevant concordance lines, despite meticulous and systematic searches. As shown in Table 6.4, when the slots were zero and one, the numbers of the whole concordance lines were eight and thirteen respectively and the number of relevant concordance lines for the associations of *large* and *damage* were six and eight respectively. However, when slots of two or three were used, the number of relevant concordance lines significantly decreased (both zero), which means that among the whole concordance lines (sixteen and eighteen respectively), there is no case where *large* is used as a pre-modifier for *damage(s)*.

Table 6.4 The frequency of the associations of *large* and *damage* with different slots in the BoE

The number of ‘slots’	The total number of concordance lines	The number of relevant lines	Percentage
No slot e.g. <i>large damage(s)</i>	8	6	75.0%
1 slot e.g. <i>large (1 word) damage(s)</i>	13	8	61.5%
2 slots e.g. <i>large (2 words) damage(s)</i>	16	0	0.0%
3 slots e.g. <i>large (3 words) damage(s)</i>	18	0	0.0%

BNCweb provides supplementary information about the collocations of dimensional adjectives. The interface of *BNCweb* is very effective in showing the collocates of a given lexical item, e.g., the collocates of a specific noun, sorted by word classes or statistical measures. Table 6.5 displays the top ten pre-modifying adjectives of the noun *damage*, sorted

by *t*-score. This shows that native speakers tend to use ‘serious’, ‘severe’ or ‘extensive’ in order to modify the noun *damage*.

Table 6.5 The top ten pre-modifying adjectives of the noun *damage* by *t*-score

No.	Word	Total No. in whole BNC	Expected collocate frequency	Observed collocate frequency	<i>t</i> -score value
1	any	119,066	53.969	357	16.04
2	environmental	8,398	3.807	137	11.38
3	serious	12,065	5.469	137	11.24
4	severe	4,550	2.062	104	10.00
5	criminal	4,329	1.962	99	9.76
6	accidental	716	0.325	72	8.45
7	extensive	4,055	1.838	73	8.33
8	much	51,955	23.550	105	7.95
9	liable	2,225	1.009	65	7.94
10	special	21,735	9.852	79	7.78

Since there may be no clear-cut norm against which to judge the collocational errors of dimensional adjectives, it is necessary to set a cut-off point to detect the deviant expressions.

In this study, the cut-off frequency to label a ‘lexical error’ is ten occurrences in the BoE. In other words, if the frequency of a specific ‘dimensional adjective + noun’ collocation drawn from the YELC is under ten occurrences in the BoE, then it is regarded as a dimensional adjective error. Although this threshold might seem arbitrary, the practice of determining a threshold as such is widely adopted in other studies with regard to collocation (e.g.

Nesselhauf, 2005; Laufer & Waldman, 2011; Ferraro, Nazar, Alonso Ramos & Wanner, 2014; Hong et al., 2011).

The BoE provides two different statistical measures for collocation: MI (Mutual Information) score and *t*-score. Each formula reflects in its own particular way the extent to which words co-occur. The advantage of statistical measures is that they can show whether a particular collocation is statistically significant or not. This study adopted *t*-score to see the collocational strength: a *t*-score of 2.00 or higher can usually be taken to be statistically significant (Hunston, 2002). It is believed that MI score is strongly affected by the raw frequency; therefore, the results based on MI score tend to be misleading (Walker, 2008; McEnery et al., 2006), particularly when the raw frequency of given lexical items is too low. In this study MI score is not used, because the raw frequency of dimensional adjectives is relatively low.

6.3.2 Research Procedures: Over-passivisation Errors of Non-alternating Unaccusative Verbs

The research procedure to show interlingual and intralingual influences on over-passivisation errors of non-alternating unaccusative verbs comprised three steps: choosing relevant sets of verbs, detecting errors and statistical analysis.

As mentioned in Section 4.2.3, this study uses six non-alternating unaccusative verbs: three verbs that can be ‘matched’ with Korean equivalents and three that cannot, as shown in Table 6.6. Although non-alternating unaccusative verbs in English can only be used as intransitives, some of their Korean equivalents can be passivised with causative/inchoative morphemes. I call the verbs that can carry the causative/inchoative morpheme and so can be passivised,

‘mismatched’ verbs. On the other hand, the verbs that do not have morphemes and so, like the English equivalents, cannot be passivised, are called ‘matched’ verbs.

Table 6.6 Matched and mismatched non-alternating verbs

	English (non-alternating verbs)	Korean (active equivalent)	Korean (causative counterpart)
Morphology	N/A	N/A	<i>-i-, -k--, -sikhi-</i>
Matched	<i>exist</i>	<i>concay-hata</i>	N/A
	<i>happen</i>	<i>ilena-ta</i>	N/A
	<i>arrive</i>	<i>tochak-hata</i>	N/A
Mismatched	<i>occur</i>	<i>palsayng-hata</i>	<i>palsayng-sikhi-ta</i>
	<i>remain</i>	<i>nam-ta</i>	<i>nam-ki-ta</i>
	<i>appear</i>	<i>nathana-ta</i>	<i>natahna-i-ta</i>

WordSmith Tools 5.0 was utilised to retrieve all the concordance lines for the six non-alternating verbs, which were then coded for error counting. We would expect that Korean learners of English would make more passivisation errors with mismatched non-alternating verbs than with matched verbs due to the L1 morphological (interlingual) influence of Korean causative counterparts.

Regarding the subject animacy effect (intralingual influence), the concordance lines were again coded according to subject animacy. Then, the numbers of errors with animate or inanimate subjects were statistically analysed. If subject animacy played a role, Korean learners would make more errors with inanimate subjects than with animate subjects in the acquisition of non-alternating unaccusative verbs.

The statistical tests were conducted through an online corpus frequency test wizard. The website¹⁹ provides some utilities from the project Statistical Inference: A Gentle Introduction for Linguists (SIGIL) (Hoffmann, Evert, Smith, Lee & Berglund-Prytz, 2008). SIGIL offers chi-square test (and G test, when needed) results between two samples. The chi-square test is widely used to determine a significant difference between the expected and the observed frequencies in one or more categories. Alternatively, the G test is used when sample sizes are very small. SIGIL automatically provides either chi-square test or G test according to the relevancy of the data. It is also worth noting that since each subject could contribute multiple data points, the independence assumption of the chi-square test is technically violated. However, because most students contributed only one data point and very few contributed as many as two or three data points, this violation is not severe and the test yields highly significant results that can be considered accurate.

6.3.3 Research Procedures: Derivational Morphological Errors

The research procedures to detect derivational morphological errors and to speculate as to the interlingual or intralingual influence on them followed the four steps presented in Figure 6.1.

In Step 1, the affixes to be extracted from the YELC for analysis were selected with reference to Bauer & Nation's (1993) list and *MorphoQuantics*.

¹⁹ It can be accessed at <http://sigil.collocations.de/wizard.html>.

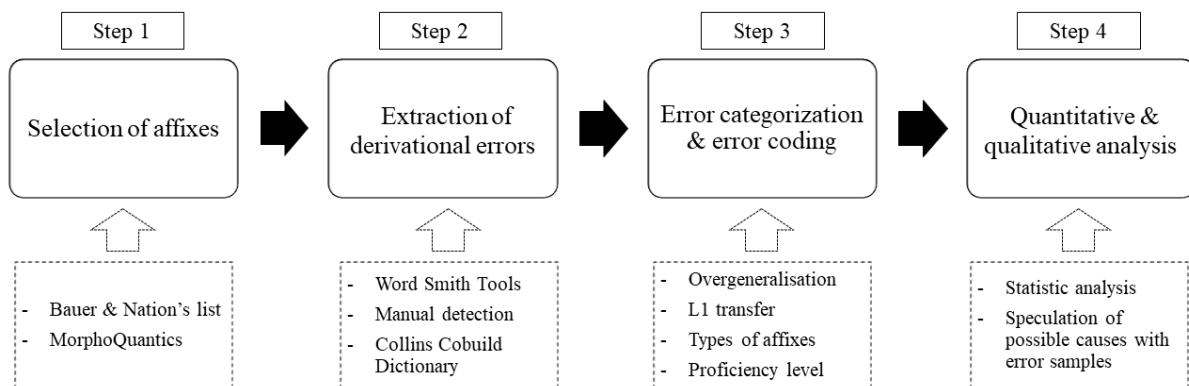


Figure 6.1 The research procedures for analysing derivational morphological errors

First, Bauer & Nation's list contains numerous English affixes. The authors set seven affixation levels using the criteria of frequency, regularity, productivity and predictability. Although these levels were established for practical reasons and have a limited theoretical basis, they offer a helpful guideline for the current study. Table 6.7 shows the levels of affixes, their descriptions and some examples (Bauer & Nation, 1993).

Table 6.7 The levels of affixes based on Bauer & Nation (1993)²⁰

Level	Description	Affixes
3	The most frequent and regular derivational affixes	-able, -er, -ish, -less, -ly, etc.
4	Frequent, orthographically regular affixes	-al, -ation, -ess, -ful, in-, etc.
5	Regular but infrequent affixes	-age, -dom, anti-, ex-, etc.
6	Frequent but irregular affixes	-ee, -ic, -ist, pre-, re-, etc.
7	Classical roots and affixes	ab-, ad-, com-, de-, dis-, etc.
-	Further affixes	cis-, semi-, pseudo-, etc.

²⁰ There are no derivational affixes available at Levels 1 and 2, because at Level 1, it is assumed that learners will not even recognise that *book* and *books* are members of the same word family, and Level 2 deals with inflection.

Secondly, as mentioned in Section 6.2.3, the current study also utilised *MorphoQuantics* (Laws & Ryder, 2014). *MorphoQuantics* divides derivational affixes into two kinds according to the position of a given root to which the affix is attached: prefix and suffix, also known as word-initial and word-final affixes respectively. Word-initial affixes (prefixes) are attached to the beginning of a word, adding information, e.g., repetition (*re-* in *replay*) or negation (*un-* in *unkind*). They are also called class-maintaining derivational affixes in that they form new words without altering the grammatical class. The word-final affixes (suffixes) are attached to the ending of a word in order to form new words by altering the grammatical class, e.g. *-er* in *teacher*, *-ness* in *happiness*, in some cases adding new information (e.g., *-able*, *-less*). Thus, they are class-changing derivational affixes (Koda, 2000).

In the selection of affixes to be investigated, *MorphoQuantics* also provides very useful information about the etymology of these affixes. For example, a suffix *-ly* that converts the word class to adverb is different from another type of *-ly* that converts the word to adjective, e.g. the *-ly* of *finally* (adverb) is different from the *-ly* of *friendly* (adjective). Similarly, according to *MorphoQuantics*, a prefix *in-* can be divided into three different kinds according to the added meaning: ‘not’ (e.g. *inconvenient*, *inactive*); ‘in, into or toward’ (e.g. *innate*, *install*); ‘of motion or direction’ (e.g. *inboard*, *inland*). This study investigated specific types of each affix (i.e. the *-ly* of adverbs and the *in-* of negation) (see Table 6.8), and *MorphoQuantics* is useful because it also provides the wordlist of each kind.

Furthermore, when selecting the relevant affixes for the analysis the token frequency of given affixes in the YELC was also considered, because the results of the study would not be robust if the tokens of certain affixes in the learner corpus were too low.

Considering all the information mentioned above, this study finally selected fifteen affixes (eight class-changing and seven class-maintaining affixes) as presented in Table 6.8.

Table 6.8 Derivational affixes extracted from the YELC

	Derivational affix	Tokens (YELC)	Added information and/or word class
Class-changing (Suffix)	<i>-able</i>	1,138	Capable of, adjective
	<i>-ible</i>	734	Capable of, adjective
	<i>-er</i>	12,892	Person or thing, noun
	<i>-less</i>	336	Without or lacking, adjective
	<i>-ish</i> ²¹	175	Adjective
	<i>-ly</i>	11,612	Adverb
	<i>-ness</i>	631	Noun
	<i>-ize(is)</i>	679	Verb
Class-maintaining (Prefix)	<i>un-</i>	1,266	Negation
	<i>in-</i>	383	Negation
	<i>im-</i>	175	Negation
	<i>il-</i>	127	Negation
	<i>ir-</i>	50	Negation
	<i>dis-</i>	1,973	Negation
	<i>re-</i>	329	Repetition

In Step 2 (see Figure 6.1) all the concordance lines with the given affixes were extracted, with the help of *WordSmith Tools 5.0*. Because the wildcard function was applied, some inappropriate cases were extracted and it was therefore necessary to ‘clean up’ the concordance lines. At this stage manual detection was employed and reference was made to Collins Online English Dictionary to identify the existence of such words. For example, in the case of words with the class-maintaining affix *im-* attached, the word *impose* was also extracted by *WordSmith Tools*. However, this should be excluded because the initial *im-* of

²¹ The original token frequency of the words with *-ish* was a lot higher than 175, but the majority of them were the word ‘English’, and therefore excluded from the analysis. Because the word ‘English’ seems to be extremely frequent, L2 learners are highly likely to acquire the word as a stand-alone item rather than a complex word (see Bauer & Nation, 1993).

impose is different from the prefix *im-* that this research is looking for. More importantly, there is an issue regarding spelling errors surrounding regular/irregular formation, or morphological transparency (Koda, 2000). That is, learners are more likely to make spelling errors with complex words that have irregular, rather than regular, formation, even though they choose the correct affix. Therefore, this research did not regard a simple spelling mistake as a derivational morphological error, because the study is concerned only with learners' choice of affixes and their cognitive process in the choice rather than the orthographical mistakes when writing essays. For example, the word **surly* (<*surely*) is definitely a spelling mistake; hence it was not counted as incorrect, in that the learner made a correct choice of the suffix (*-ly*). However, these kinds of spelling errors are addressed in the next section, which deals with spelling errors in the phonological/orthographic domain.

In Steps 3 and 4 (see Figure 6.1), the data were analysed both qualitatively and quantitatively and an attempt was made to categorise some types of derivational morphological errors by the degree of overgeneralisation or L1 transfer. In addition, the acquisition of derivational affixes (morphological awareness) was investigated according to the types of affixes or the proficiency level. The online corpus frequency test wizard SIGIL (see Section 6.3.2) was also used for statistical tests.

6.3.4 Research Procedures: Orthographic (Spelling) Errors

In order to efficiently detect a large quantity of spelling errors in the YELC, this study followed three steps, utilising *WordSmith Tools*. Unlike the other three error features mentioned above, where I used the Concord tool in *Wordsmith Tools* to detect the target forms, in the case of spelling errors I mainly utilised the Wordlist tool.

First, three Wordlist files, one for each proficiency level, were created in order to obtain the frequency numbers for each individual word type. These Wordlist files contain not only spelling errors but also correctly spelled words. In order to delete all the correctly spelled words and have only the spelling errors, the Match function from Wordlist was employed. With this function, one can delete the entries that match with a ‘Wordlist match list’ file that is uploaded separately, in this case the Text file of the BNC world corpus.²² Since the BNC world corpus contains almost 100 million words with over 500,000 types, more than the number of English dictionary entries, it is very useful to remove correctly spelled words from the Wordlist files.

Secondly, after correctly spelled words were deleted, it was necessary to manually clean the Wordlist files, because these still contained many entries that needed to be removed, for example numbers, proper nouns and newly coined words that do not yet appear in the BNC world corpus, such as *facebook*, *iphone*, *website* and *MP3*. Then, as mentioned in Section 5.2.3, the entries made by incorrect morphological selections were also deleted.

Finally, the remaining entries of spelling errors in the Wordlist files were also manually investigated and separately coded in order to count the frequencies of errors in each category. Note that this study did not consider all spelling errors, but focused only on specific kinds: three pairs of consonant confusions ([p]-[f], [l]-[r], [b]-[v]) for interlingual errors; inflectional and derivational spellings for intralingual errors. A number of spelling errors were categorised as interlingual or intralingual errors, but the others, which might be larger in number, remained uninvestigated and so uncategorised. This shows how difficult it is to specify the underlying sources of spelling errors and to categorise them.

²² The Wordlist file of the BNC world corpus can be freely downloaded from the website www.lexically.net/downloads/version4/downloading%20BNC.htm. This Wordlist file was converted into a Text file in order for it to be uploaded as a ‘Wordlist match list’ file in *WordSmith Tools*.

Furthermore, in order to calculate error occurrence rates more accurately, this study examined the frequencies of correctly spelled words for each spelling error. It first found the target form (correct form) of each error and then used *WordSmith Tools* to detect how frequently the correct forms occur in the YELC. The comparison between the frequencies of errors and correct forms is expected to show more accurate error occurrence rates for both interlingual and intralingual spelling errors. In addition, this study analysed spelling errors according to learners' proficiency levels in order to show whether these errors become fossilised with higher English proficiency. As for statistical tests, the SIGIL (see Section 6.3.2) was also used.

6.4 Chapter Summary

This chapter has presented the two main research questions and sub-questions to be answered by the corpus-based error analysis which are used to show interlingual and intralingual influences on the four error features, one from each lexical domain. It has also introduced the corpora used for the error analysis, as well as the research techniques utilised for each error feature. Finally, it has described the research procedures for each of the four error features in more detail.

Although the research methods and procedures for each error feature were slightly different, in each case the analysis followed the same basic processes. First, this study made a careful choice of relevant target items in order to show clearly the interlingual and intralingual influences. Second, in each case *WordSmith Tools* was utilised to search for the target items in the YELC. Reference corpora and statistical measures were used as and when appropriate

to reveal the error sources for each error feature. During these processes, several research techniques and programs were used to assist the corpus-based error analysis.

CHAPTER 7: COLLOCATIONAL ERRORS OF DIMENSIONAL ADJECTIVES

Chapters 7 to 10 explore four kinds of lexical errors, one for each lexical domain, in order to demonstrate the grounds for the new error taxonomy based on the L2 lexical development model. First, with regard to the semantic domain, this chapter focuses on collocational errors of dimensional adjectives.

In Section 7.1, I outline the errors with reference to six dimensional adjectives, with examples drawn from the YELC. Then, in Sections 7.2 and 7.3, I show negative and positive interlingual influence (L1 transfer) respectively as a possible source of the errors, with corresponding examples. In Section 7.4, I address intralingual influence as another possible source, with some examples detected by means of the corpus-based error analysis.

7.1 Outline of the Corpus-based Error Analysis for Collocational Errors of Dimensional Adjectives

WordSmith Tools extracted a total of 1,640 tokens of the six dimensional adjectives from the YELC. The frequency of each adjective is shown in Table 7.1. Through comparison with the reference corpora, this study found 127 collocational errors of dimensional adjectives, which is about 7.7% of the total frequency of the six dimensional adjectives in the YELC. The fifth and sixth columns respectively show the number of texts in which errors occur and the text occurrence rate. These figures indicate how collocational errors are dispersed in terms of variability. If the rate were too low, we could claim that the data were skewed, which would

suggest that a small number of writers very frequently made the same errors while the majority of writers did not produce the errors. If this were the case, then we could not claim that the findings reflect the characteristics of the data in a relevant manner. However, since the overall text occurrence rate is quite high (88.9%), the data can be regarded as valid.

Table 7.1 The frequency of collocational errors of dimensional adjectives in the YELC²³

Dimensional adjectives	Error frequency	Token frequency ²⁴	Error percentage ²⁵ (%)	Number of texts in which errors occur	Text occurrence rate ²⁶ (%)
<i>large</i>	19	92	20.7%	17	89.5%
<i>small</i>	59	340	17.4%	51	86.4%
<i>high</i>	18	410 ²⁷	4.4%	17	94.4%
<i>low</i>	24	166	14.5%	20	83.3%
<i>long</i>	3	439	0.7%	3	100%
<i>short</i>	4	193	2.1%	4	100%
Total	127	1,640	7.7%	112	88.9%

Collocational errors, that is, deviant associations of ‘dimensional adjective + noun’, are typically represented in two forms: 1) a restrictive form (adjective + noun); 2) a predicative form (noun + linking verb + adjective).

²³ More information according to proficiency level is provided in Appendix 3.

²⁴ The token frequency excludes occurrences where such items are not used as adjectives. It also excludes items used as idiomatic expressions, e.g. *in short*, *by and large*.

²⁵ The error percentage is presented only for additional information of descriptive statistics in order to outline the collocational errors of dimensional adjectives, not for inferential statistics. The error percentage here would not carry any statistical meaning in this analysis, because the token frequency would be more dependent upon registers or topics.

²⁶ The formula to obtain text occurrence rate is: Number of texts in which errors occur / Error frequency * 100.

²⁷ The collocations ‘high school’ and ‘high way’ were excluded from the total frequency, because these seem to be more like compound words rather than the associations of ‘pre-modifier (adjective) + noun’ form.

- 1) The restrictive form is where a dimensional adjective located in front of a noun directly pre-modifies the noun, e.g., “some teachers give children large punishment” (Sentence from Text file 2893_02).
- 2) The predicative form is where a noun located in a subject position is described by a dimensional adjective with a linking verb *BE* or *GET*, e.g., “[...] military should be large [...]” (Sentence from Text file 1867_02).

Table 7.2 gives several examples of deviant associations of dimensional adjectives with nouns. Some of the collocational errors seem to be influenced by learners’ L1 (Korean). For example, associations such as *large damage* and *small fight* do not appear deviant when literally translated into Korean, because in Korean the noun *damage* (손해) can be modified by the adjective *large* (넓), according to the Sejong Corpus (see Section 6.2.4). Similarly, the Korean equivalent of the association *small fight* (작은싸움) is commonly used in the Korean language. It could be argued that these kinds of collocational errors are evidence of interlingual influence (or L1 transfer) in the acquisition of dimensional adjectives.

Table 7.2 Examples of collocational errors of dimensional adjectives (deviant associations)

Dimensional adjectives	Deviant nouns associated with adjectives
<i>large</i>	accident, convenience, damage, disadvantage, effort, etc.
<i>small</i>	attention, battle, confidence, experience, fight, habit, memory, etc.
<i>high</i>	difficulty, effect, loyalty, personality, violation, etc.
<i>low</i>	accessibility, authority, dignity, effectiveness, health, etc.
<i>long</i>	castle, discipline, waterfall.
<i>short</i>	opinion, pain, punishment, cigarette.

On the other hand, there are some associations that are not common at all, even in Korean, e.g., *small confidence*, *high personality*, *low health*. These examples show that the L2 acquisition of dimensional adjectives could be affected not only by L1 transfer, but also by another source, that is, intralingual influence (the example concordance lines of the collocational errors are provided in Appendix 3). Table 7.3 and Figure 7.1 show the occurrence rate of the collocational errors of each dimensional adjective according to error source. As can be seen in the table and figure, Korean learners appear to make many more intralingual errors than interlingual errors (86.6% vs. 13.4%). This suggests that the lexical representations of those dimensional adjectives mostly overlap between the two languages (Korean and English), although there remains a relatively small area that belongs to only one of the two languages (see Section 4.1). However, this finding should be interpreted with caution, because the total frequency of each dimensional adjective is quite low.

Table 7.3 The error occurrence rates of collocational errors of dimensional adjectives according to error source

Dimensional adjectives	Interlingual errors	Intralingual errors
<i>large</i>	6 (31.6%)	13 (68.4%)
<i>small</i>	5 (8.5%)	54 (91.5%)
<i>high</i>	1 (5.6%)	17 (94.4%)
<i>low</i>	4 (16.7%)	20 (83.3%)
<i>long</i>	1 (33.3%)	2 (66.7%)
<i>short</i>	0 (0.0%)	4 (100%)
Total	17 (13.4%)	110 (86.6%)

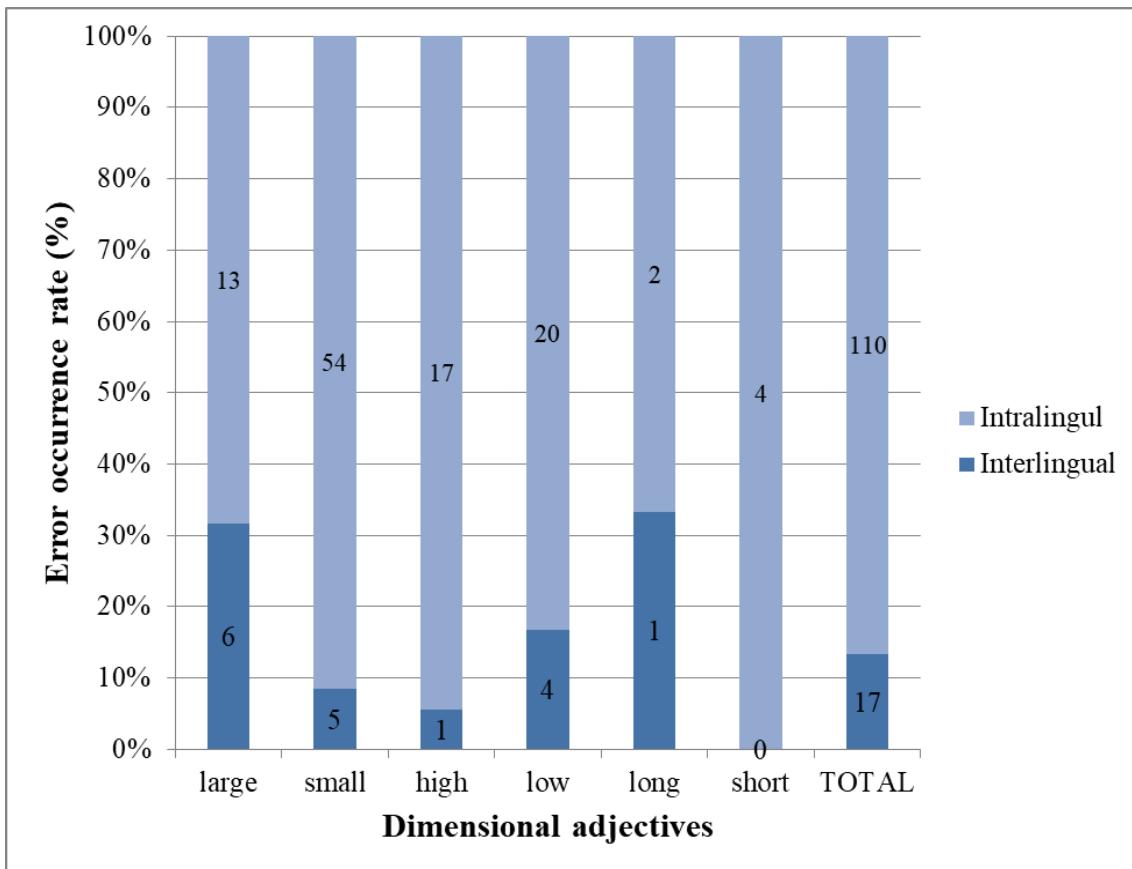


Figure 7.1 The error occurrence rates of collocational errors of dimensional adjectives

Sections 7.2 to 7.4 provide actual examples from naturally occurring data and discuss the possible sources of the dimensional adjective errors.

7.2 Interlingual Influence (L1 Transfer): A Possible Source of Collocational Errors of Dimensional Adjectives

It has been largely accepted that interlingual influence is one of the main factors that can cause errors in L2 acquisition (Jiang, 2000; Odlin, 1989; Ringbom, 1987, 2007) and the case

of collocational errors would be no exception. This section shows that interlingual influence plays a significant role when Korean learners of English make collocational errors of dimensional adjectives. To do so, it presents four illustrative examples from the YELC (i.e., *large accident*, *small (and low) attention*, *high loyalty* and *fidelity* and *low accomplishment*) and shows how these collocational errors are affected by L1 transfer.

The first deviant collocation I would like to look at is *large accident*, as in (1). A writer might produce the deviant association *large accident*, where he or she intends to express the idea of a ‘serious’ or ‘fatal’ accident. According to the BNC, the adjectives most frequently used to pre-modify *accident* are ‘fatal’ and ‘serious’. However, the Korean translation of *large accident*, 大사고, is quite a common collocation according to the Sejong Corpus.

- (1) Even if many driver abide a law, if a driver lost his or her mind, it would lead a large accidents and make many good drivers be killed too. ([sic], Sentence from Text file 1626_02)

In order to check the validity of the deviant collocation, first, I searched the BoE for the kinds of adjectives that usually pre-modify the noun *accident(s)*. *LookUp* software extracted 4,276 concordance lines for ‘adjective + *accident(s)*’. Figure 7.2 below shows the results of the picture function for the noun *accident(s)* in the BoE, with its collocates in each column sorted by raw frequency. It also shows that the five most frequent adjectives are ‘fatal’, ‘serious’, ‘tragic’, ‘nuclear’ and ‘freak’, when highlighting the pre-modifiers in the NODE-1 slot.

in	a	fatal	NODE	and	the	the
of	the	serious	NODE	in	a	a
was	of	tragic	NODE	the	in	in
a	an	nuclear	NODE	<p>	and	to
the	and	freak	NODE	at	branch	and
after	in	general	NODE	inquiry	to	of
to	s	industrial	NODE	that	was	s
by	this	terrible	NODE	of	</h1>	was
had	by	more	NODE	but	is	that
been	worst	happy	NODE	on	it	<p>
s	that	major	NODE	or	of	he
have	some	marine	NODE	he	that	at
for	from	unfortunat	NODE	which	he	it
that	to	horrific	NODE	i	said	is
as	no	minor	NODE	investigat	<p>	on
and	two	bad	NODE	involving	had	<dt>
is	for	such	NODE	to	which	be
<p>	very	historical	NODE	it	i	had
be	any	nasty	NODE	as	as	<bl>
just	with	other	NODE	a	his	by
at	most	medical	NODE	is	emergency	i
or	have	little	NODE	when	on	have
number	or	separate	NODE	are	than	as

Figure 7.2 The picture function for *accident(s)* in the BoE with its collocates in each column sorted by raw frequency

Table 7.4 illustrates these five most significant adjectives that pre-modify the noun *accident(s)* sorted by *t*-score. The *t*-scores of the five pre-modifying adjectives are all over 14.00, which means these collocations do not occur coincidentally. On the other hand, the raw frequency of the collocation ‘*large + accident(s)*’ in the BoE is only 1 and its *t*-score is 1.00. Therefore, the collocation *large accident(s)*, shown in the YELC as in (1), can be regarded as a collocational error of the dimensional adjective *large* (see Section 6.3.1 for more information about the cut-off point of *t*-score).

Table 7.4 The five most significant pre-modifying adjectives of *accident(s)* sorted by *t*-score

Pre-modifying adjective (Node-1 collocate)	Raw frequency of the collocation	<i>t</i> -score	NODE
fatal	430	20.74	<i>accident(s)</i>
serious	246	15.65	
tragic	216	14.70	
freak	211	14.53	
nuclear	211	14.50	

The second examples show that dimensional adjectives are associated with the noun *attention*, as presented in (2) and (3) below. According to the BNC, the noun *attention* is frequently modified by ‘little’ or ‘scant’, rather than ‘small’ or ‘low’ in a similar sense.

- (2) Then, in case of drivers do the multitasking and pay just small attention to driving, the possibility of accident grows. ([sic], Sentence from Text file 1330_02)
- (3) Therefore, drivers who are apt to drive may not have difficulty about low attention during driving. ([sic], Sentence from Text file 888_02)

Following the same procedure as used for the analysis of *accident*, I searched the BoE for the collocation ‘adjective + *attention(s)*’ and extracted 12,941 concordance lines. The five most significant pre-modifying adjectives of *attention(s)* are presented in Table 7.5. The *t*-scores of these five adjectives are all over 22.00, which means the associations are very significant, and thus they can be considered as standard English collocations. On the other hand, the collocation *small attention(s)* occurs only once in the BoE and its *t*-score is 1.00. The raw frequency of the collocation *low attention(s)* in the BoE is 4 and its *t*-score is 1.99. In both

cases, the *t*-score is below 2.00; hence, the collocations are regarded as significantly underused associations of ‘adjective + noun’ (Hunston, 2002). Because the deviant collocations *small attention* and *low attention* never or rarely occur in the BNC or the BoE, they can be considered as collocational errors of dimensional adjectives. Interestingly, the Korean translations of *small attention* (작은 관심) and *low attention* (낮은 관심) are both common expressions, according to the Sejong Corpus. Therefore, these examples may be further evidence of interlingual influence (L1 transfer) in the L2 acquisition of dimensional adjectives.

Table 7.5 The five most significant pre-modifying adjectives of *attention(s)* sorted by *t*-score

Pre-modifying adjective (Node-1 collocate)	Raw frequency of the collocation	<i>t</i> -score	NODE
more	1433	37.12	<i>attention(s)</i>
much	1380	36.86	
little	674	25.70	
special	662	25.63	
public	516	22.47	

Regarding the third examples, it seems that some collocational errors of dimensional adjectives are produced owing to confusion caused by other L2 words that have a similar meaning. In the example sentences (4) and (5) below, the writer(s) used the nouns *loyalty* and *fidelity* associated with the adjective *high*.

- (4) I think salary solider service is the best solution. Then loyalty for country will be higher. ([sic], Sentences from Text file 52_02)

- (5) I agree that physical punishment in army make soldiers have high fidelity and royalty.
([sic], Sentence from Text file 645_02)

According to the BoE, the abstract noun *loyalty* is rarely associated with *high* (only 1 occurrence). The association ‘*high + fidelity*’ can also be regarded as a deviant collocation, because although the noun, *fidelity*, is frequently modified by the dimensional adjective *high* (109 occurrences), the collocation *high fidelity* is used as a technical term related to the sound quality in stereo recording, but not with the meaning intended as in (5). Hence, both *high loyalty* and *high fidelity* can be regarded as collocational errors. As shown in example (5), some learners do use ‘*fidelity*’. There is a slight difference in meaning between *fidelity* and *loyalty*, but as mentioned, native speakers use ‘*high fidelity*’ only in a specific context related to sound quality. However, in Korean, the abstract nouns *fidelity* and *loyalty* can be both translated by the same word, *충성도*. Consequently, the translations of *high loyalty* and *high fidelity* are acceptable and correct in Korean. These examples show how the L1 can cause learners to select an incorrect collocate. Learners might struggle to choose the correct collocation in the mental lexicon.

A search of the BoE for ‘adjective + *loyalty (loyalties)*’ yielded 1,595 occurrences. Table 7.6 shows the five most significant adjectives of *loyalty (loyalties)*, i.e. *political, fierce, great, tribal* and *total*. Since the *t*-scores are all over 2.00, these five collocations are statistically significant and so can be regarded as acceptable combinations, as opposed to a dimensional adjective error such as *high loyalty (loyalties)*. In other words, the Korean writer should have

used, for example, *great loyalty* in order to deliver a similar sense of *high loyalty*, according to Table 7.6.

Table 7.6 The five most significant pre-modifying adjectives of *loyalty (loyalties)* sorted by *t-score*

Pre-modifying adjective (Node-1 collocate)	Raw frequency of the collocation	<i>t-score</i>	NODE
political	79	8.83	<i>loyalty (loyalties)</i>
fierce	46	6.78	
great	47	6.75	
tribal	41	6.40	
total	36	5.97	

The fourth example sentences (6) and (7) below are a slightly different case from that of *high loyalty* and *high fidelity*. The collocation *low accomplishment(s)* cannot be found in either the BoE or the BNC. Instead, the data from reference corpora show that *low achievement* is a more frequent combination. The raw frequency of the collocation *low achievement* is 23 and the *t-score* is 4.79, according to the BoE. Therefore, it could be argued that example (6) is a collocational error of the dimensional adjective *low*, while example (7) is not a collocational error. Interestingly, since both *accomplishment* and *achievement* can be commonly translated by the same Korean noun 성취, Korean learners could be confused when selecting the appropriate collocate for the dimensional adjective *low*.

- (6) When students begin to hate teacher who punished them, it can cause low academic accomplishment. ([sic], Sentence from Text file 855_02)

(7) So young children have possible to show low educational achievement. ([sic],
Sentence from Text file 2197_02)

Data from the BoE show that the noun *accomplishment(s)* is very frequently associated with the dimensional adjective *small*, rather than *low*, as shown in Figure 7.3. The raw frequency of the collocation *small accomplishment(s)* is 9 and the *t*-score is 2.91, based on the BoE²⁸. Interestingly, the Korean translation of the collocation *small accomplishment(s)* (*작은 성취*) is commonly used according to the Sejong Corpus. Therefore, it might be challenging for L2 learners to discern which expressions are semantically congruent and can therefore be literally translated from the L1 into the L2.

1 will attest, that is no small accomplishment. The Bosnia decision:
2 is to translate this small accomplishment into a peace settlement.
3 November 1918, that was no small accomplishment. Yet the internal
4 homosexuality in France--no small accomplishment in a country that to
5 to the electoral lists--no small accomplishment. The name of this
6 the peace, which is no small accomplishment. His handling of Europe
7 Effenberg, which is no small accomplishment, and if Craig had to
8 and keep a daily list of small accomplishments. You'll be amazed how
9 were most needed. No small accomplishments these, brought on by a

Figure 7.3 The concordance lines of the collocation *small accomplishment(s)* in the BoE

The above pair of examples produced by Korean learners, i.e. (6) and (7), suggest that, as regards the semantic incongruity of learners' mental lexicon, interlingual influence in the L2 acquisition of collocations of dimensional adjectives is not always negative, but can also be positive. Examples of positive L1 transfer are presented in the next section.

²⁸ Note that among the 9 concordance lines in Figure 7.3 the majority (7 cases) are *no small accomplishment(s)*. While there may be particular reasons why the collocation *small accomplishment(s)* is frequently associated with the collocate *no*, this is not within the scope of the current research.

7.3 Positive Interlingual Influence (Positive L1 Transfer)

The previous section presented evidence of the interlingual influence (L1 transfer) that plays a significant role in producing learners' collocational errors of dimensional adjectives. As has been noted, some errors seem to be caused by semantic incongruity between L1 (Korean) and L2 (English) (see Section 3.3.3). Semantic incongruity occurs when some of the associations in a language do not perfectly overlap with those in other languages. In these cases, L2 learners may not be able to avoid collocational errors, because they may instinctively resort to their L1 semantic information and thus produce inappropriate collocations (see Section 3.5.1). However, if the L1 has its own lexically conventionalised expressions, which overlap with those in the target language, L2 learners' transfer of L1 semantic information to L2 expressions will have a positive outcome. This can be applied to collocations of dimensional adjectives. Presumably, the number of collocations where there is positive L1 transfer is greater than that related to negative L1 transfer, because humans perceive the same real-life dimensions in the world.

This section illustrates a couple of examples of positive L1 transfer in the YELC related to the nouns *difficulty* and *punishment*, the two nouns that will be featured in the next section (Section 7.4) with regard to collocational errors caused by intralingual influence. It is worth looking at positive L1 transfer in these cases, because learners might also make collocational errors with the same nouns.

The first example of positive L1 transfer is related to the noun *difficulty*. As shown in Figure 7.4, the collocation *great difficulty* occurs 715 times in the BoE, with the *t*-score 26.73, which means the collocation is statistically significant.

10 his huge army had great difficulty subduing a force of merely 1000
11 in particular had great difficulty with this concept, as well as with
12 themselves had great difficulty staying awake! On the other hand
13 includes words of great difficulty, e.g. import-export, comedy
14 plant it. `We have great difficulty getting anything to grow.
15 companies have great difficulty raising money for product

Figure 7.4 Some of the concordance lines of the collocation *great difficulty* in the BoE

Likewise, the collocation *great difficulty* is also commonly used in Korean (큰 어려움), according to the Sejong Corpus. It could be argued that this is an example of positive L1 transfer, because this collocation can be found in the YELC, as in (8).

- (8) There are millions of Internet sites today and trying to change all the peoples' ID would take enormous time and money. It would give the Internet users great difficulty.
([sic], Sentences from Text file 2319_02)

However, interestingly, a Korean learner makes a deviant association, *high difficulty* (see (14) in Section 7.4). The data from the BoE show that the noun *difficulty* is frequently associated with other dimensional adjectives, *little*, *big* and *great*, rather than *high*.

The other example is a collocation related to the abstract noun *punishment*. This noun is frequently associated with the dimensional adjective *heavy*, which denotes the concept of weight: *heavy* appears in the list of the 50 most frequent adjectives in the BNC and the BoE, sorted by *t*-score. The collocation *heavy punishment(s)* occurs 23 times in the BoE, with the *t*-score 4.80. Some concordance lines are presented in Figure 7.5.

16 seamer Fraser suffered heavy punishment with figures of 1-63,
17 run off -- but Luther took a heavy punishment for his vengeance. <o>
18 become anxious if we received heavy punishment in attacking the
19 the time Rumanians - taking heavy punishment from the German machine
20 Darren Gough coming in for heavy punishment, Mullally's
21 attempt, voters inflicted heavy punishment on the ruling party.
22 the government side is taking heavy punishment, losing hundreds of men

Figure 7.5 The concordance lines of the collocation *heavy punishment(s)* in the BoE

The noun *punishment(s)* can also be pre-modified by the dimensional adjective *heavy* in Korean (*무거운*). This suggests that L2 learners' mental lexicon shares the same or similar conceptualisation of some dimensional adjectives, which results in positive L1 transfer. Indeed, a few cases can be found in the YELC, as illustrated in (9) and (10).

(9) Therefore to prevent heavy corporal punishment in schools, Korean wrong education curriculums which are only want student to study needs to be changed. ([sic], Sentence from Text file 3178_02)

(10) There are not just a way to ease the punishment for students because the rules in the penalty card system were very strict and when students got several penalty cards there were heavy punishment such as calling parents to come school, or if worse, get expelled. ([sic], Sentences from Text file 1742_02)

Once again, however, the YELC contains a case in which a Korean learner has made a collocational error, *low punishment* (see (16) in Section 7.4), an example that shows how the same abstract noun, 'punishment', can be associated not only with an appropriate adjective, but also with an inappropriate adjective.

The effect of positive L1 transfer can be seen in (8) to (10), examples of cases where some of the most frequent collocations in the BoE also occur in the L2 learners' interlanguage. The

fact that the L1 transfer can result in both positive and negative influences indicates that using L1 information in the L2 acquisition can be a double-edged sword. It may be very helpful to L2 learners, if positive L1 transfer is promoted and negative L1 transfer is avoided in the L2 vocabulary acquisition. The caveat is that L2 learners might not be able to recognise which collocations are acceptable and which ones are not when they use literal translation from L1 into L2. Consequently, they would benefit from explicit instruction to heighten their awareness of collocations of dimensional adjectives.

7.4 Intralingual Influence: Another Possible Source of Collocational Errors of Dimensional Adjectives

This section of the thesis explores evidence in the YELC of another possible source of collocational errors of dimensional adjectives, intralingual influence; that is, influence from the target language (see Section 3.4.2), rather than the interlingual influence. L2 learners confront many situations in which they cannot be helped by L1 semantic information and have to find relevant words despite the lack of L2 vocabulary in their mental lexicon. In this case, they are likely to make an ‘approximate selection’ within the limited range of their L2 vocabulary. This section illustrates four pairs of examples of collocational errors of dimensional adjectives that show intralingual influence.

The first pair of examples are related to the dimensional adjective *small*.

- (11) I had not good memorizing ability. I forgot almost things in 1 month. So I have one or two very small memories. ([sic], Sentences from Text file 671_01)

In example (11), the writer seems to have struggled to find a relevant dimensional adjective in their mental lexicon. Note that the collocation *small memory* occurs only 3 times in the BoE. Given that the *t*-score is only 1.73, this collocation could be regarded as a collocational error of the dimensional adjective *small*. Since there is no such collocation in the Sejong Corpus (i.e. *작은 기억* in Korean), this error is probably not due to interlingual influence. Instead, it seems more reasonable to conclude that this error results from intralingual influence, through approximate selection. We can only speculate as to the meaning the writer intended to express. It might be a ‘dim memory’ or ‘vague recollection’. According to the BoE, the noun *memory* is pre-modified most frequently by such adjectives as ‘fond’, ‘happy’, ‘good’, ‘recent’ or ‘distant’.

Similarly, the deviant collocation *small confidence(s)*²⁹ as illustrated in (12) does not occur in the BNC, which means that this collocation can be regarded as a collocational error of the dimensional adjective *small*. Data from the BoE show that the abstract noun *confidence* is frequently pre-modified by such adjectives as ‘public’, ‘great’ or ‘complete’, but not ‘small’. There seems to be no interlingual influence in this case, because the Korean translation (i.e. *작은 확신*) of the collocation does not occur in the Sejong Corpus.

- (12) Some people may think that using a cell phone during driving a car is time-efficient and not dangerous. However, one small confidence can make terrible accident. ([sic], Sentences from Text file 1229_02)

²⁹ This collocation occurs 6 times in the BoE. Based only on the *t*-score, 2.4486, it might not be regarded statistically as an error. However, taking into account the cut-off point of the raw frequency set up in this study (under ten occurrences), it can be considered an error (see Section 6.3.1 for more information about the cut-off score). Therefore, the collocation *small confidence(s)* can be seen as on the borderline and we should always note that this is not a clear cut matter; rather, there is a continuum of appropriateness.

One possible explanation for the deviant collocations in (11) and (12) is that the writers might assume that the abstract nouns *memory* and *confidence* are measurable in terms of volume. These collocations may sound acceptable to native speakers, if they can be considered as figurative expressions. However, even if considered as figurative, they would still not be within a range of standard English (see Section 4.1.3). In the production of figurative expressions speakers' cognitive processes, referred to in this study as 'cognitive vehicles' (Section 3.4.2), play a vital role. In other words, no matter what labels are attached to them, these collocations seem to be produced in the same way as other deviant expressions and they may be collocational errors, if their meanings are overextended.

Interestingly, in the BoE the abstract noun *confidence(s)* is frequently associated with other dimensional adjectives, like *high* and *low*. The collocation *high confidence(s)* occurs 33 times, with the *t*-score 5.30 and the collocation *low confidence(s)* occurs 15 times, with the *t*-score 3.61. Figure 7.6 includes some of the concordance lines of *high* (or *low*) *confidence(s)* in the BoE. This suggests that English native speakers recognise this abstract noun *confidence* with the dimensional concept of height, rather than volume.

23 asserts that the US can have 'high confidence' in the safety
24 woes. Only 9 percent expressed high confidence in Wall Street. The
25 because I think they have a very high confidence that none of us, from
26 Neill (Marine Corp # we have a high confidence factor that, in fact,
27 badly that anyone should have high confidence of an easy victory. I
28 negotiations, and you get a very high confidence level between

29 stress disorder; social anxiety; low confidence; examination and
30 he suffered from self-doubt and low confidence in his first season
31 Because I've always had a very low confidence level in my physical
32 few people gave them a high or low confidence rating. One possible
33 but before I'd got into a rut of low confidence. Once that happens you
34 chances. It seems they have low confidence. I've been in that

Figure 7.6 The concordance lines of the collocation *high* (and *low*) *confidence(s)* in the BoE

The second pair of examples are related to the dimensional adjective *high*, as shown in (13) and (14).

- (13) It gonna be hard to critize the person who have a high effect on society and web the information that has a bad effect on some big company. ([sic], Sentence from Text file 230_02)
- (14) Without proper places, smokers cannot enjoy their smoking as they want. They hide and hide to smoke, and finally the difficulty of controlling the smokers gets higher. ([sic], Sentences from Text file 1751_02)

The collocations in these two examples, i.e. *high effect* (높은 효과), *high difficulty* (높은 어려움), are regarded as collocational errors because they are not commonly used either in English or in Korean. According to the BoE, the collocation *high effect* does not occur at all and *high difficulty* occurs only once. When I searched for dimensional adjectives that frequently modify the abstract nouns *effect* and *difficulty*, I found ‘little’, ‘big’ and ‘great’, among the top 50 collocates, sorted by *t*-score. However, these three dimensional adjectives denote a different dimensional concept (amount or size) and not the concept of height (see Section 7.3). In other words, it could be argued that English native speakers conceptualise the abstract nouns *effect* and *difficulty* as something that can be measured by a scale of amount or size, whereas some Korean learners of English consider the nouns to be measurable by height. Indeed, these errors do not seem to be caused by the influence of the L1, because the Korean equivalents of these collocations cannot be found in the Sejong Corpus. Consequently, it could also be argued that in producing these figurative overextensions the writers were using cognitive vehicles with L2 sources.

The third pair of examples are collocational errors associated with another dimensional adjective, *low*: *low accessibility* and *low punishment*, as in (15) and (16).

- (15) U.S and Canada has very restricted smoking area. Low accessibility to smoking place can lower the smoking rate. ([sic], Sentences from Text file 1299_02)
- (16) But setting the standards about corporal punishment is totally necessary. it must be low punishment and it cannot give children psychological hurts. ([sic], Sentences from Text file 3178_02)

These deviant associations seem to be made in a similar manner to those of the dimensional adjective *high*, in that the writers might assume that these abstract nouns can be measured by a scale of height. Once again, these errors are not affected by interlingual influence, because the literally translated equivalents of *low accessibility* (낮은 접근성) and *low punishment* (낮은 벌) do not occur in the Sejong Corpus. Neither do they occur in the BoE (zero occurrences for both collocations).

Similarly, one can assume that the Korean learners who produced the overextensions *short cigarette* and *long waterfall*, as in (17) and (18) respectively, were conceptualising *cigarette* and *waterfall* in terms of measurement scales. Neither of these collocations occurs in the BoE, or in the Sejong Corpus.

- (17) Passing people would not get a good image of the owners or representatives of the buildings with short cigarettes on the floor. ([sic], Sentence from Text file 1735_02)
- (18) In there, grand forests, long and wide waterfall, kind rural people, and anything I love exist but Korean cultures. ([sic], Sentence from Text file 900_01)

In summary, the collocational errors of dimensional adjectives shown in this section seem to reflect intralingual influence, whereby cognitive vehicles lead L2 learners to make collocational errors with the L2 sources.

As cognitive vehicles are defined in this study as processes of human mental activities that are universal among all language users (see Section 3.4.2), similar types of intralingual errors to those shown in this section can also be made by young L1 speakers, as well as other L2 learners who have different L1 backgrounds. Among the studies that have explored this phenomenon, Ravn & Gelman (1984) conducted an experiment which showed that young children (3- to 5-years-old) were confused between size and height and that native speaker children often believed that *big* means *tall* (see also Maratsos, 1973, 1974). Furthermore, Maloney & Gelman (1987) showed that even L1 undergraduate native speakers could interpret the dimensional adjective *big* in different settings as size or height. These studies seem to support the argument that the particular errors produced by Korean learners and discussed in this section clearly show intralingual influence rather than interlingual influence.

7.5 Conclusion

This chapter has illustrated several collocational errors of dimensional adjectives in the semantic domain. The collocational errors detected in the YELC by means of corpus-based error analysis show that Korean learners of English make collocational errors of dimensional adjectives that are affected by both interlingual and intralingual influences, as discussed with reference to the proposed new error taxonomy (see Section 3.5.2).

For example, deviant expressions such as *small attention* and *low attention* could be regarded as collocational errors because they rarely, or never, occur in the reference corpora. Since the

Korean translations of these expressions are in common usage, according to the Sejong Corpus, it appears that the errors are affected by learners' L1. However, interlingual influence is not invariably negative; it can also act in a positive way, as in the collocations *great difficulty* and *heavy punishment*. The findings thus suggest that interlingual influence is not always an obstacle to the L2 acquisition, but can also be helpful to L2 learners to develop vocabulary, if those learners are guided by language teachers to raise their awareness.

On the other hand, deviant expressions such as *small confidence* and *high effect*, which can also be regarded as collocational errors, seem to be affected by intralingual influence with the action of cognitive vehicles, because these expressions do not occur in the BoE or BNC, or even in the Sejong Corpus. Instead, they appear to be created by learners in a logical way with the help of cognitive vehicles and limited L2 sources.

Therefore, it is concluded that L2 learners' can be affected by both interlingual and intralingual influences, which can result in the production of collocational errors of dimensional adjectives in the semantic domain.

CHAPTER 8: OVER-PASSIVISATION ERRORS OF NON- ALTERNATING UNACCUSATIVE VERBS

This chapter explores one error feature in the syntactic domain, namely over-passivisation errors of non-alternating unaccusative verbs, in order to show how interlingual and intralingual influences can play roles in the L2 acquisition of the grammatical features in the L2 mental lexicon.

In Section 8.1, I introduce the three types of over-passivisation errors of non-alternating unaccusative verbs that occur in the learner corpus data. In Sections 8.2 to 8.4, I investigate the interlingual influence of L1 morphology on over-passivisation errors, by illustrating the errors detected in the YELC and comparing error occurrence rates between matched and mismatched verbs. This includes examining whether the interlingual influence becomes fossilised according to learners' proficiency levels and checking whether the findings on interlingual errors might be skewed by individual verb influence. In Sections 8.5 to 8.7, I address the intralingual influence, by investigating how subject animacy affects over-passivisation errors. As in the investigation of the interlingual influence, in these sections I also explore fossilisation according to proficiency level and individual verb influence.

8.1 Three Types of Over-passivisation Errors of Non-alternating Unaccusative Verbs

Corpus-based error analysis was carried out to investigate the behaviours of six non-alternating unaccusative verbs (three 'matched' and three 'mismatched' verbs: see Section 6.3.2 for more information). The analysis found that the over-passivisation errors of these

non-alternating unaccusative verbs were represented in three different ways. The different kinds of error representations would not necessarily have been considered if this study had followed the great majority of previous studies by conducting a grammaticality judgement test, in which it is possible to control the conditions of test items (see Section 4.2.2). However, this study deals with learner corpus data, in which learners could produce any kind of syntactic form with the six verbs investigated here. Indeed, all three kinds of errors identified can be considered as over-passivisation errors, because learners violated the passive rules in terms of ‘unaccusativity’ in the use of non-alternating unaccusative verbs. I classify the three types³⁰ of over-passivisation errors in the YELC as Type 1 (passive error); Type 2 (‘BE + base form’); and Type 3 (transitive error carrying an object), as follows:

(1) Type 1: typical passive errors

- a. *These possible problems were already happened many times. ([sic], Sentence from Text file 2974_02)
- b. *Then, second-hand smoking is occured. ([sic], Sentence from Text file 1035_02)

First, Type 1 is the prototypical kind of over-passivisation error, as learners produce passive forms with non-alternating unaccusative verbs. Because they are considered to be the most frequent type of over-passivisation error, they have been used by previous studies as the items of grammaticality judgement tests in order to check whether subjects make such errors. Example sentences of Type 1 errors are drawn from the YELC, as in (1). The passivisation of the verbs *happen* and *occur* in (1a) and (1b) respectively is not grammatically correct.

³⁰ These exclude such cases as misspellings and tense errors of the six verbs, which are not directly related to the over-passivisation of non-alternating unaccusative verbs.

(2) Type 2: BE + base form

- a. *For example, very critical disease is appear. ([sic], Sentence from Text file 1937_02)
- b. *... the disadvantage of going to army is exist. ([sic], Sentence from Text file 3124_02)

Type 2 over-passivisation error is ‘BE + base form’, as exemplified in (2). In the sentences in (2a) and (2b), the verbs *appear* and *exist* should not be used with the linking verb *is*. There are two possible readings for this error type: 1) The learner intended to use the passivised form as in Type 1 errors but mistakenly omitted the morpheme *-(e)d*; 2) The learner incorrectly identified the verb as an adjective, so thought it should be accompanied by a linking verb. If reading 1) is the case, these errors can be regarded as Type 1 over-passivisation errors of non-alternating unaccusative verbs. However, if the learner mistakenly considered the verb as an adjective as in reading 2), then this should not be classified as an over-passivisation error. In order to be absolutely certain, it might be necessary to query the writers as to their intention when writing the phrases, a task that would be almost impossible and certainly undesirable, given that the YELC was established in 2011 (Rhee & Jung, 2014).

This study provisionally follows reading 1) and assumes that the writers intended to make passivised forms of the verbs, so that Type 2 error is considered to belong to the group of over-passivisation errors. However, since reading 2) is also plausible, this study conducted a statistical analysis with the error frequency excluding Type 2 (see Table 8.2 in Section 8.2). If the writers did take the verbs for adjectives as in reading 2), it would be interesting to investigate the possible causes of this phenomenon. However, that investigation is not within the scope of the current study.

(3) Type 3: transitive error carrying an object

- a. *In addition, when I arrive the top of mountain, ... ([sic], Sentence from Text file 2880_01)
- b. *I want to remain good relationship with her forever. ([sic], Sentence from Text file 2462_01)

Type 3 over-passivisation error, termed ‘transitive error carrying an object’, occurs when non-alternating unaccusative verbs are incorrectly used as transitive verbs, carrying objects, as in the sentences in (3a) and (3b). In (3a), the verb *arrive* incorrectly carries the objective phrase ‘the top of mountain’, and similarly, in (3b) the verb *remain* is incorrectly accompanied by the adjectival phrase ‘good relationship’ as an object. These can undoubtedly be considered as over-passivisation errors, because the learners seem to have recognised the verbs as transitive verbs and accordingly believed that they could use passivised forms of them. Thus, it seems that when making Type 3 errors, the learners are following the same process that is involved in Type 1 errors.

In short, the three types of error representations described above are all regarded as over-passivisation errors with non-alternating unaccusative verbs, because all three types involve the violation of so-called ‘unaccusativity’ (more example concordance lines of the unaccusative verb errors are also provided in Appendix 4).

8.2 Interlingual Influence: L1 Morphology

As mentioned in Section 6.3.2, I used *WordSmith Tools* to extract 872 occurrences of the matched verbs and 453 occurrences of the mismatched verbs; these totals include not just

correct forms, but also all three types of over-passivisation errors. Table 8.1 shows the total and error frequencies of matched and mismatched non-alternating unaccusative verbs in the YELC including Type 2 errors (BE + base form), while Table 8.2 shows the frequencies without Type 2.

Table 8.1 The distribution of the matched and mismatched non-alternating verbs including the cases of Type 2 (BE + base form)

	Matched	Mismatched
Total frequency	872	453
Error form frequency	141	162
Error percent (%)	16.2%	35.8%

As shown in Table 8.1, learners made 141 errors with matched non-alternating unaccusative verbs, representing 16.2% of the total frequency of those verbs in the YELC; and 162 errors with mismatched verbs, representing 35.8% of the total frequency. To compare the numbers of correct form and error form frequency between matched and mismatched non-alternating unaccusative verbs, a chi-square test was carried out. The result demonstrates the statistically significant difference between matched and mismatched verbs (*chi-square* = 63.771, *p value* < .001).

Table 8.2 The distribution of the matched and mismatched non-alternating verbs excluding the cases of Type 2 (BE + base form)

	Matched	Mismatched
Total frequency	872	453
Error form frequency	83	148
Error percent (%)	10.2%	33.7%

Table 8.2 shows the frequencies without Type 2 errors. Once again, the chi-square test indicates that the errors of the two groups (matched and mismatched) are significantly different ($\text{chi-square} = 103.33$, $p \text{ value} < .001$). Therefore, even if Type 2 (BE + base form) is not considered as an over-passivisation error, it appears that the interlingual influence on over-passivisation errors of non-alternating unaccusative verbs is still plausible. In fact, the difference in error percentage between matched and mismatched verbs is greater in that case than when including Type 2 errors (see Tables 8.1 and 8.2: 16.2% - 35.8% vs. 10.2% - 33.7%), so that the explanatory power of the statistical measure for the interlingual influence is even greater. As mentioned in Section 8.1, this study assumes that Type 2 errors can also be regarded as over-passivisation errors. Accordingly, in the sections that follow, the error frequencies of the over-passivisation errors include all three types presented in Section 8.1.

The example sentences in (4) below contain over-passivisation errors of three matched verbs: *exist*, *happen* and *arrive*. As mentioned in Sections 4.2.3 and 6.3.2, the Korean equivalents of the matched verbs cannot carry causative morphemes and so there are no causative counterparts in the Korean language. That is to say, the English matched verbs and their Korean equivalents behave alike in terms of passivisation: passivised forms are regarded as errors.

(4) Errors of matched verbs

- a. *... but if proper physical punishment is not existed, school is not controlled students and school will raise the bad human. ([sic], Sentence from Text file 84_02)
- b. *And, one year later, in this vacation, the most sad thing in my life was happened. ([sic], Sentence from Text file 2085_01)
- c. *As soon as we arrive the hotel, we firstly unpacked our bags. ([sic], Sentence from Text file 1021_01)

For example, the passivised form ‘is existed’ in (4a) is a passivisation error both in English and in Korean, which indicates that there is no L1 transfer by means of literal translation. Hence, it could be argued that the over-passivisation errors with these verbs produced by Korean learners as in (4) are not caused by the L1 (Korean). Rather, the errors in (4) appear to result from incorrect application of passivisation rules, which could be considered intralingual influence.

On the other hand, the example sentences in (5) show over-passivisation errors of three mismatched verbs: *occur*, *remain* and *appear*. Unlike the case of the matched verbs, the Korean equivalents of these three mismatched verbs do have causative counterparts that carry corresponding morphemes (see Table 6.6 in Section 6.3.2).

(5) Errors of mismatched verbs

- a. *The best interesting thing that happened in last summer vacation was occured in my high school. ([sic], Sentence from Text file 2998_01)
- b. *I can’t understand why this stupid system is remained this 21c society. ([sic], Sentence from Text file 510_02)

- c. *As cellular phones began to be sold by almost all the public, many new-uprising problems including this problem has been appeared so often. ([sic], Sentence from Text file 752_02)

For example in (5b), the English verb *remain* cannot be passivised, but it does have a causative counterpart in Korean (*namkita* 남기다). This may confuse Korean learners of English, so that they are liable to make over-passivisation errors of mismatched verbs. Hence, it could be speculated that the influence of learners' L1 (interlingual influence) plays a vital role in producing over-passivisation errors of mismatched verbs. This does not necessarily mean that intralingual influence does not also play a part in producing these errors with mismatched verbs, because it is still possible that some of the learners were influenced by incorrect application of passivisation rules (intralingual influence) even with mismatched verbs, although it would be very hard to specify which influence affects which errors. This is why the current study conducted statistical measurements: as shown in Tables 8.1 and 8.2 above, chi-square tests revealed significant differences between the error frequencies of matched and mismatched verbs.

The findings of the current study with regard to the interlingual influence (L1 transfer) are in line with those of previous studies. However, these findings are novel and significant in that they are drawn from naturally occurring data which deal with a productive language skill, whereas previous studies have collected their data from grammaticality judgement tests that are mainly involved with a receptive skill (e.g. Chung, 2014; Kondo, 2005; Yip, 1995; Zobl, 1989). Overall, it can be argued that the interlingual influence of L1 morphology plays a significant role in the L2 acquisition of English non-alternating unaccusative verbs. Therefore,

this result could still fill a gap in the understanding of the L2 acquisition of non-alternating unaccusative verbs by offering support for the existing claim as to interlingual influence; that is, it reveals that the influence is still valid in productive language presentation, which is considered to involve a greater cognitive burden than receptive language skills.

8.3 Interlingual Influence According to Proficiency Level

In order to identify whether or not there is interlingual influence from L1 morphology at different proficiency levels, this study investigated the error frequencies of non-alternating unaccusative verbs in each sub-corpus: low, intermediate and advanced levels (see Table 6.1 in Section 6.2.1). If the over-passivisation errors are not fossilised, then the error occurrence rate should decline as the English proficiency level rises from low to advanced. Otherwise, the error rate should increase or remain constant.³¹

Table 8.3 Error occurrence rates for L1 morphology according to proficiency level

	Low	Intermediate	Advanced	Total
Matched verbs	42 / 194 (21.6%)	98 / 639 (15.3%)	1 / 39 (2.6%)	141 / 872 (16.2%)
Mismatched verbs	41 / 91 (45.1%)	118 / 341 (34.6%)	3 / 21 (14.3%)	162 / 453 (35.8%)
Note: the figures in the cells indicate: error frequency / total frequency (error percent)				

³¹ According to Han (2003: 99), ‘fossilisation’ can be defined as “stabilised interlanguage forms that remain in learner speech or writing over time, no matter what the input or what the learner does”. Strictly speaking, it might need a more synthetic judgement based on comprehensive learner data to conclude whether or not certain error features are fossilised, which might be difficult. However, the current study could show one aspect of fossilisation by analysing the comparisons between error frequencies of English proficiency level groups.

As shown in Table 8.3 and Figure 8.1, the error percentage of both matched and mismatched verbs clearly declines as the proficiency level rises from low to advanced. It is also worth noting that, just as the total error occurrence rate of the matched verbs (16.2%) is lower than that of the mismatched verbs (35.8%), so at every proficiency level the error occurrence percentage of the matched verbs is lower than that of the mismatched verbs.

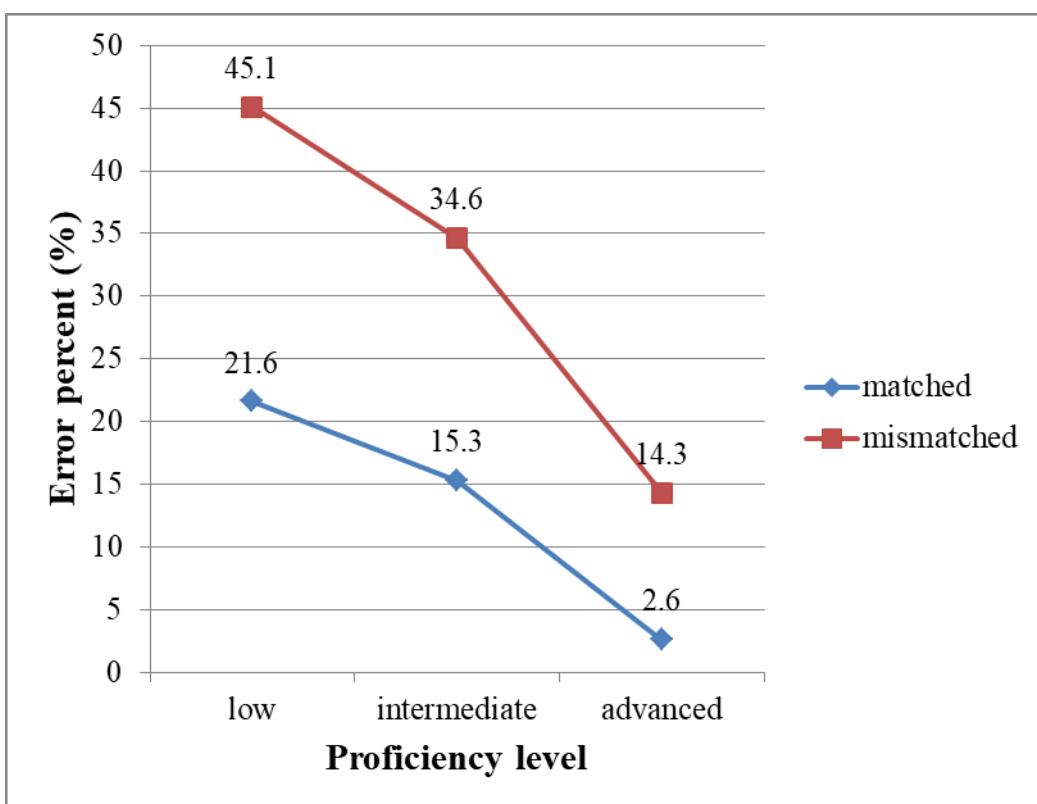


Figure 8.1 Error occurrence rates for L1 morphology by proficiency level

In order to investigate the interlingual influence of L1 morphology according to proficiency level, chi-square tests were conducted to find out whether there exist statistically significant differences between proficiency groups. The tests were carried out for both matched and mismatched verbs. As shown in Table 8.4, there are significant differences between low and

advanced groups for both matched (*p value* < .01) and mismatched verbs (*p value* < .05). Although the differences between low and intermediate and between intermediate and advanced, are not statistically significant, they are only just below significance level in the case of the matched verbs (*p value* = .05118 and .05008) and quite close to it in the case of mismatched verbs (*p value* = .08648 and .09347).

Table 8.4 Results of chi-square tests between proficiency level groups

	Matched verbs	Mismatched verbs
Low / Intermediate	3.8026 (<i>p value</i> = .05118)	2.9387 (<i>p value</i> = .08648)
Intermediate / Advanced	3.8389 (<i>p value</i> = .05008)	2.8137 (<i>p value</i> = .09347)
Low / Advanced	6.6426**	5.5439*
Note: ** <i>p value</i> < .01 * <i>p value</i> < .05		

According to these findings, the interlingual influence of L1 morphology on over-passivisation errors of non-alternating unaccusative verbs decreases as learners' proficiency level increases. It seems that the errors do not become fossilised, as learners in advanced groups make significantly fewer errors than learners in low groups. These findings are in line with those of Chung's (2015) longitudinal study, which showed evidence of Korean learners' improvement in the L2 acquisition of unaccusative verbs. In short, the interlingual influence of L1 morphology becomes weaker as learners' proficiency in English increases.

8.4 Interlingual Influence of Individual Verbs

This section addresses the error frequencies of each individual non-alternating unaccusative verb. If the error rates for one or two idiosyncratic verbs were extraordinarily high, this could skew the results as to the statistical significance of differences in error occurrences between matched and mismatched verbs. In that case, the findings reported in Sections 8.2 and 8.3 could be misleading and could not be used as evidence that over-passivisation errors are caused by interlingual influence.

Table 8.5 Error occurrence rates of each individual verb

	Verb	Error frequency	Total frequency	Error rate (%)	Number of texts in which errors occur	Text occurrence rate³² (%)
Matched verbs (+)	<i>exist</i>	55	214	25.7%	49	89.1%
	<i>happen</i>	78	614	12.7%	68	87.2%
	<i>arrive</i>	8	44	18.2%	7	87.5%
Mismatched verbs (-)	<i>occur</i>	94	242	38.8%	73	77.7%
	<i>remain</i>	54	135	40.0%	54	100%
	<i>appear</i>	14	76	18.4%	13	92.9%
TOTAL		303	1,325	22.9%	264	87.1%

Table 8.5 shows the error occurrence rate of each individual verb explored in this study. The sixth and seventh columns respectively show the number of texts in which errors occur and the text occurrence rate. Similar to the collocational errors of dimensional adjectives in the previous chapter (see Table 7.1 in Section 7.1), these figures show how unaccusative verb errors are dispersed. As the overall occurrence rate is over 87.1%, the data indicate that the

³² The formula to obtain the text occurrence rate is: Number of texts that contain errors / Error frequency * 100.

unaccusative errors occur in different texts, which suggests that the data are not skewed by a small number of idiosyncratic writers. The information from other columns is presented in a different form in Figure 8.2, where the length of the bars indicates the error occurrence rate of each verb. The plus symbol (+) is used to indicate matched verbs, whereas the minus symbol (-) indicates mismatched verbs. As can be seen from the table and figure, in almost every case the individual matched verbs have lower error rates than any of the mismatched verbs (see Appendix 4 for the error percentage for each individual verb according to proficiency level).

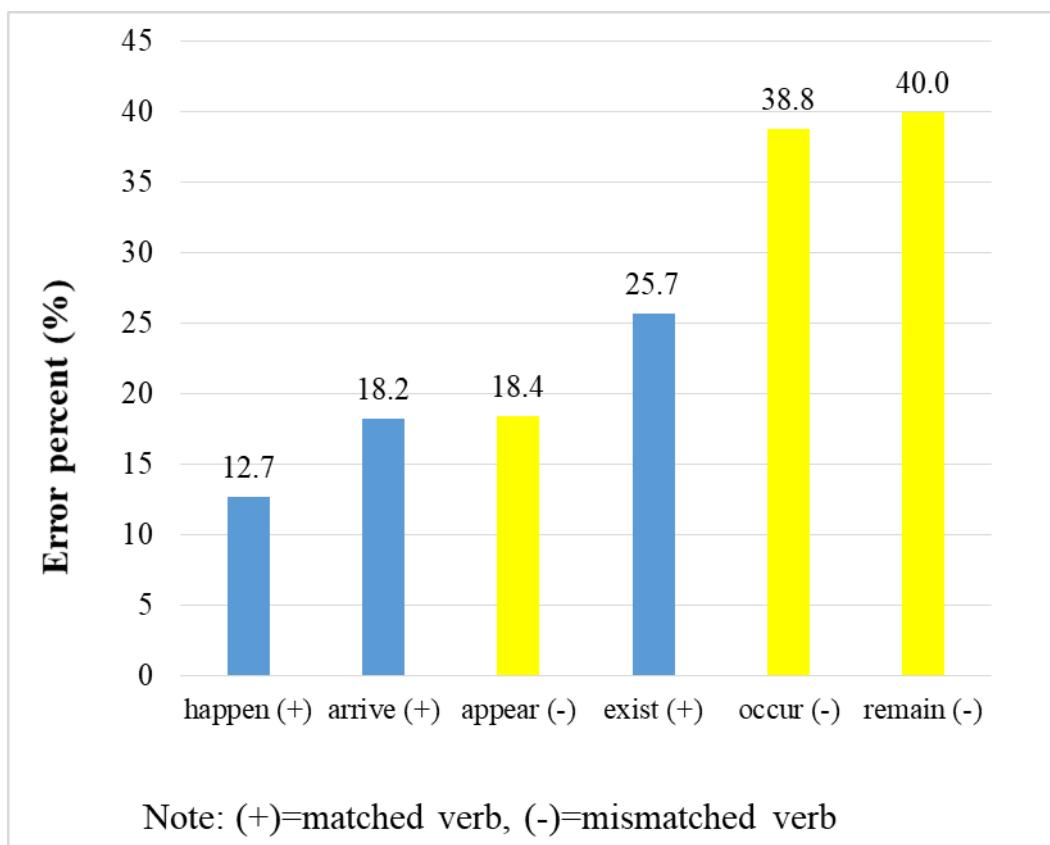


Figure 8.2 Error occurrence rates of each individual verb

However, there is one exception. The mismatched verb *appear* has a lower error rate (18.4%) than the matched verb *exist* (25.7%) (or, *exist* has a higher rate than *appear*). *Appear* has the

lowest over-passivisation error rate among the mismatched verbs, with 14 errors out of 76 occurrences in the YELC.

It is not immediately clear why *appear* (or *exist*) is an outlier in these results. One possible reason may be related to L2 input, albeit that the impact of the L2 input seems to be limited. This claim is based on the assumption that the less frequently L2 learners are exposed to a non-alternating unaccusative verb, the more likely they are to make over-passivisation errors with that verb. However, the relationship between the error occurrence rates and the amount of L2 input is not within the scope of the current study. Another possible reason for the outlier, *appear* (or *exist*), could be related to the subject animacy effect, which will be discussed in the next section. However, neither of these two possibilities seems to give a clear explanation for this peculiar case.

Overall, however, the investigation into error occurrence rates of individual verbs shows that, despite the one outlier, there is no significant individual verb effect. Consequently, these results do not rebut the evidence presented in Section 8.2 of interlingual influence on over-passivisation errors of non-alternating unaccusative verbs.

In order to clarify the relevancy of the interlingual influence of L1 morphology without the frequency of the verb *appear* (or *exist*), another statistical tests were conducted. The results show significant differences ($G\text{-square} = 74.51220$, $p\ value < .001$ without *appear* and $\chi^2 = 78.36930$, $p\ value < .001$ without *exist*) between matched and mismatched verbs.

8.5 Intralingual Influence: Subject Animacy Effect

In Sections 8.2 to 8.4, I made reference to the interlingual influence on over-passivisation errors of matched and mismatched verbs. In this section, I seek to clearly identify the role

played by intralingual influence in producing over-passivisation errors of the same non-alternating unaccusative verbs, by investigating the subject animacy effect. As mentioned in Section 4.2.4, the subject animacy effect in the use of unaccusative verbs could be regarded as deriving from one of the L2 (English) sources of over-passivisation errors. Therefore, in order to show how L2 learners over-passivise non-alternating unaccusative verbs according to the subject animacy effect, I counted the numbers of error forms and correct forms of non-alternating unaccusative verbs, according to whether or not the verbs adopt animate subjects.

(6) Errors with animate subjects

- a. *Although people must use their real name on the Internet, **many people** can be appeared by ways such IP searching. ([sic], Sentence from Text file 3015_02)
- b. *So if animals should not be used in medical experiments, **animals** used to eating should not be exist. ([sic], Sentence from Text file 611_02)

(7) Errors with inanimate subjects

- a. *If urgent **situation** will be happened, men should protect our South Korea. ([sic], Sentence from Text file 3006_02)
- b. *I think that **internet** must be remained as anonymous. ([sic], Sentence from Text file 2888_02)

The example sentences (6) and (7) drawn from the YELC illustrate errors with animate and inanimate subjects respectively. In (6a) and (6b), the animate subjects ‘many people’ and ‘animals’ are used with incorrect passivised forms of non-alternating unaccusative verbs, whereas in (7a) and (7b) the inanimate subjects ‘situation’ and ‘internet’ are used with incorrect passivised forms.

Indeed, it has already been noted in the literature that language users normally prefer to have animate subjects with transitive verbs and unergative verbs. In contrast, it seems natural for language users to tend to adopt inanimate subjects for non-alternating unaccusative verbs, as they comply with the canonical mapping rule for those verbs (Hartshorne, Pogue & Snedeker, 2015). Native English speakers probably recognise (or unconsciously know) the subjects of the non-alternating unaccusative verbs as having the thematic role of Theme, whether they adopt animate or inanimate subjects. However, as mentioned in Section 4.2.4, the reason why L2 learners make over-passivisation errors, as in (6) and (7), might be because they recognise the subjects as having the thematic role of Agent rather than Theme. If the subject animacy (the intralingual influence) does play a role in over-passivisation errors, we could predict that L2 learners will make more errors when they adopt animate subjects, because animate subjects have more volitional power than inanimate subjects.

Table 8.6 Error occurrence rates for subject animacy

	Total frequency	Error form frequency	Error percent (%)
Animate subject	137	40	29.2%
Inanimate subject	1,188	263	22.1%
Total	1,325	303	22.9%

Table 8.6 shows error occurrence rates of the non-alternating unaccusative verbs according to subject animacy. Clearly, Korean learners prefer to have inanimate subjects (1,188 occurrences) with non-alternating unaccusative verbs, rather than animate subjects (137 occurrences). As hypothesised, Korean learners make slightly more errors with animate subjects (29.2%) than with inanimate subjects (22.1%).

According to the results of a chi-square test, the difference in error occurrences between the two groups (animate subject vs. inanimate subject) is not significant (*chi-square* = 3.0815, *p value* = 0.062). However, as the *p value* is only just above significance level (.05), the subject animacy effect on the over-passivisation errors is still regarded as very relevant.

Table 8.7 Error occurrence rates according to L1 morphology and subject animacy

	Subject animacy	Error frequency	Total frequency	Error rate (%)
Matched verbs	Animate	18	87	20.7%
	Inanimate	123	785	15.7%
	Sum	141	872	16.2%
Mismatched verbs	Animate	22	50	44.0%
	Inanimate	140	403	34.7%
	Sum	162	453	35.8%

To be clear, it is worth investigating whether the error occurrence rates shown in Table 8.6 could also be affected by the interlingual influence of L1 morphology, as the error frequencies include both matched and mismatched verbs and so the two variables (subject animacy and L1 morphology) might work together, or in conflict with one another. As reported in Table 8.7, the error occurrence rates with animate subjects in both matched (20.7%) and mismatched verbs (44.0%) are greater than those with inanimate subjects (15.7% and 34.7% respectively), which is in line with the total error rates in Table 8.6. Again, according to chi-square tests, for both matched and mismatched verbs the differences in error occurrences between animate and inanimate subject groups are not statistically significant, which is also similar to the result for the total error rates.

Interestingly, Chung (2014) and Pae et al. (2014) also found significant differences in over-passivisation errors between animate and inanimate subjects. However, in contrast to the current study, in which Korean learners of English make more over-passivisation errors with animate subjects, the Korean learners in Chung's (2014) and Pae et al.'s (2014) studies made significantly more errors with inanimate subjects than with animate subjects.

One possible reason might be related to the nature of the data. As already noted, both Chung (2014) and Pae et al. (2014) used a grammaticality judgement test that forced students to answer yes or no for each task item. In contrast, the current study examines learner corpus data, a large dataset comprising texts written by L2 learners in natural, real-world contexts. This again offers an important new perspective from which to consider the L2 acquisition of unaccusative verbs. That is, the current study extends the boundary of the research on unaccusative verbs by adding a novel type of data related to productive language skills, whereas previous studies have focused only on receptive data from grammaticality judgement tests. In other words, this study shows that Korean learners make more errors with animate subjects when they 'produce' English, whereas when they 'comprehend' English, as in Chung (2014) and Pae et al. (2014), they make more errors with inanimate subjects.

In summary, it can be argued that subject animacy has some effect in the L2 acquisition of non-alternating unaccusative verbs, albeit not to an extent that is statistically significant.

8.6 Intralingual Influence According to Proficiency Level

Following the same procedure as for interlingual influence (see Section 8.3), this study also investigated the error frequencies for subject animacy in each sub-corpus in order to identify whether or not the errors caused by intralingual influence become fossilised according to

proficiency level. As shown in Table 8.8 and Figure 8.3, the error percentage of both groups (animate and inanimate subjects) clearly declines as the proficiency level rises from low to advanced.

Table 8.8 Error occurrence rates for subject animacy according to proficiency level

	Low	Intermediate	Advanced	Total
Animate subjects	13 / 27 (48.1%)	27 / 103 (26.2%)	0 / 7 (0.0%)	40 / 137 (29.2%)
Inanimate subjects	70 / 258 (27.1%)	189 / 877 (21.6%)	4 / 53 (7.5%)	263 / 1188 (22.1%)
Note: the figures in the cells indicate: error frequency / total frequency (error percent)				

The error rate for animate subjects is higher than at the low and intermediate levels. However, the error occurrence rate for inanimate subjects becomes higher than that for animate subjects at the advanced level. This might be because there is a significantly low frequency (only 7 occurrences with 0 errors) of animate subjects at the advanced level, hence the impact might be limited. The results of chi-square tests between proficiency groups show differences that are either significant, or only just below significance level. Therefore, it could be argued that, similar to the interlingual influence of L1 morphology shown in Section 8.3, the errors caused by intralingual influence do not become fossilised.

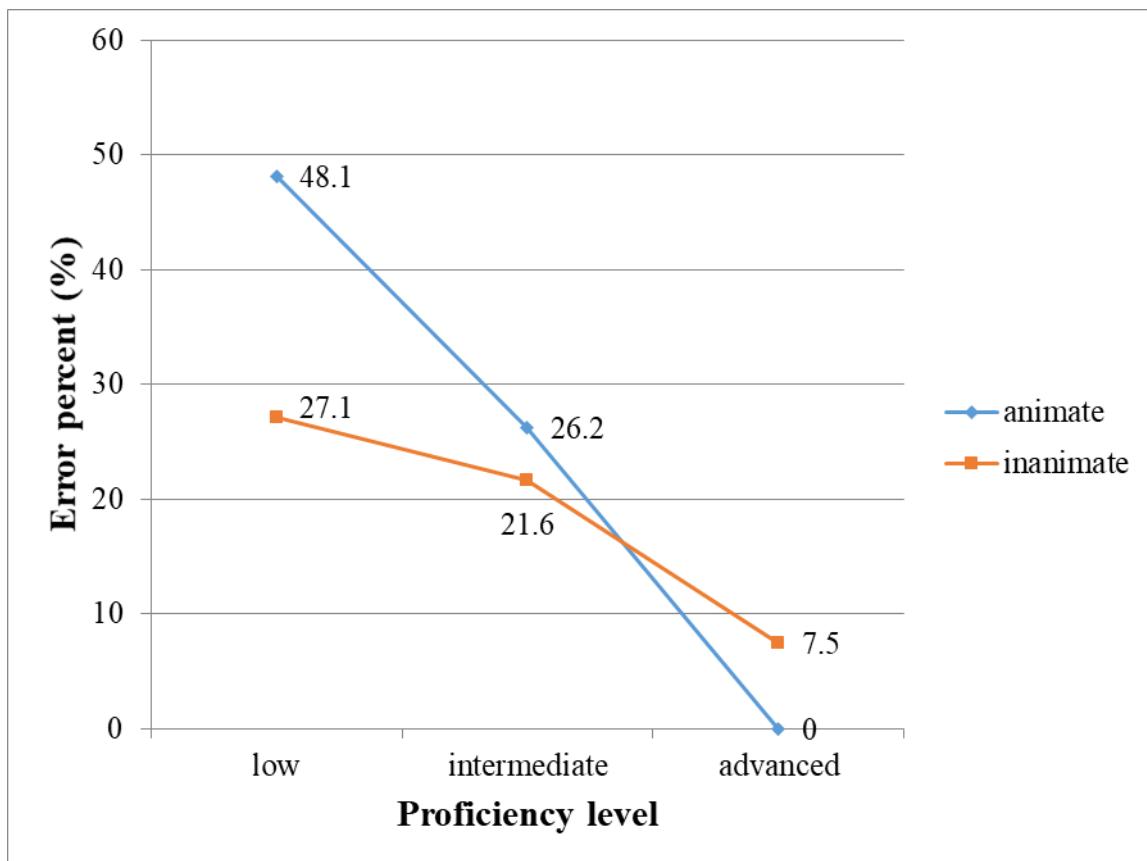


Figure 8.3 Error occurrence rates for subject animacy by proficiency level

8.7 Intralingual Influence of Individual Verbs

As Section 8.4 has shown the interlingual influence of individual verbs, it is also worth examining whether or not the results regarding the intralingual influence of subject animacy might be skewed by the effect of one or more individual verbs. In order to check the individual verb influence, Table 8.9 shows the error occurrence rate of each individual non-alternating unaccusative verb for animate and inanimate subjects, followed by the results of chi-square tests.

Table 8.9 Error occurrence rate of each individual verb and the results of chi-square test

Verb	Animate subject (error / verb frequency)	Inanimate subject (error / verb frequency)	Chi-square test
<i>exist</i>	7 / 39 (17.9%)	48 / 175 (27.4%)	not significant
<i>happen</i>	3 / 6 (50.0%)	75 / 608 (12.3%)	significant
<i>arrive</i>	8 / 42 (19.0%)	0 / 2 (0.0%)	not significant
<i>appear</i>	3 / 15 (20.0%)	11 / 61 (18.0%)	not significant
<i>remain</i>	14 / 30 (46.7%)	40 / 105 (38.1%)	not significant
<i>occur</i>	5 / 5 (100.0%)	89 / 237 (37.6%)	significant

According to the chi-square tests, although only two verbs (*happen* and *occur*) show a significant difference between error occurrences for animate and inanimate subjects, Korean learners generally make more errors with animate subjects than with inanimate subjects for all but one verb, *exist*. Note that in the analysis on interlingual influence of individual verbs (see Section 8.4), the error occurrence rate of one of the matched verbs, *exist*, was exceptionally high compared to the other two matched verbs and even one of the mismatched verbs, *appear*. According to Table 8.9, this seems to be closely related to the error rate of *exist* with inanimate subjects (27.4%), which was unexpectedly higher than that with animate subjects (17.9%). This peculiar behaviour of the verb *exist* falls outside the scope of the current study, but would be worth investigating in future research. However, since the total frequency of *exist* in the YELC is 214 occurrences, which represents about 16.2% of the total frequency of all six verbs, it seems that this outlier has only limited impact on error rates due to the subject animacy effect. Overall, it seems that the individual verb influence found in this section does not skew the results of the subject animacy effect on non-alternating unaccusative verbs indicated in Section 8.5.

8.8 Conclusion

This chapter has investigated how the over-passivisation errors of six non-alternating unaccusative verbs are affected by both interlingual and intralingual influences, based on L2 learners' naturally occurring data: a learner corpus that consists of written essays. The errors investigated here represent one of the error features in the syntactic domain.

Firstly, this study selected three non-alternating unaccusative verbs (matched verbs) that have corresponding Korean equivalents and another three verbs (mismatched verbs) that do not have Korean equivalents. The statistical measurements of the errors of matched and mismatched verbs clearly demonstrate the interlingual influence of L1 morphology by showing significantly more errors with mismatched verbs than with matched verbs. Moreover, the results according to proficiency level indicate that these errors do not become fossilised, as they significantly decrease with higher proficiency. Although further investigation revealed one outlier verb, there is no evidence that the results are skewed by individual verb influence.

Secondly, with regard to intralingual influence, the investigation has shown the effect of subject animacy on the errors. Korean learners show a clear preference to use inanimate subjects with non-alternating unaccusative verbs and yet they also made more errors with animate subjects than with inanimate subjects. Although this difference between animate and inanimate subjects is not statistically significant, the findings show that the animacy of subjects as intralingual influence could play an important role in the L2 acquisition of non-alternating unaccusative verbs. As with the interlingual influence of L1 morphology, it seems that the errors caused by the subject animacy effect do not become fossilised, since they decline with an increase in proficiency level; furthermore, individual verb influence is limited.

In addition, the findings of this chapter offer a new perspective from which to consider syntactic errors, in terms of how the comprehended syntactic knowledge can be similarly or differently produced in the L2 language representation. For example, the findings in this chapter with regard to the interlingual influence of L1 morphology are in line with previous studies, in that they show significant influence of L1 transfer on the L2 acquisition of unaccusative verbs. On the other hand, the results with regard to the intralingual influence through the subject animacy effect are partly differentiated from those of previous studies.

Overall, it is concluded that L2 learners may produce over-passivisation errors in the use of non-alternating unaccusative verbs, which can be affected by both interlingual and intralingual influences. These findings offer evidence to support the grounds for the new error taxonomy proposed in Section 3.5.2.

CHAPTER 9: DERIVATIONAL MORPHOLOGICAL ERRORS

This chapter provides evidence of interlingual and intralingual influences on the L2 acquisition of distributional knowledge of derivational morphology, based on naturally occurring data produced by Korean learners of English. Corpus-based error analysis was carried out to detect derivational morphological errors as one error feature in the morphological domain in the proposed new error taxonomy presented in Section 3.5.2.

In Section 9.1, I give an overview of the corpus-based error analysis, which reveals the derivational morphological errors according to the kinds of affixes (class-changing vs. class-maintaining affixes). In Section 9.2, I discuss whether or not these errors become fossilised, by presenting statistical measures based on the error frequencies between L2 learners' proficiency levels. In Sections 9.3 and 9.4, I propose a division of derivational morphological errors into four types according to the degree of overgeneralisation and illustrate some examples of each error type with regard to the intralingual influence. In Section 9.5, I focus on one of the four error types (Type D, see Table 9.4 in Section 9.3) and provide illustrative examples in order to reveal the interlingual influence involved in the production of derivational morphological errors.

9.1 Overview of the Corpus-based Error Analysis on Derivational Morphological Errors

As mentioned in Section 6.3.3, this study investigated L2 learners' derivational morphological errors in the use of the fifteen affixes (i.e. eight class-changing and seven

class-maintaining affixes). This section investigates whether or not L2 learners may have difficulty in the acquisition of certain kinds of affixes. It seems that L2 learners acquire certain affixes earlier than others, which may be due to the degree of difficulty of the affixes, or the extent to which the learners are exposed to those particular affixes. However, to date, despite meticulous and systematic searches, it appears that no research has used L2 learners' naturally occurring data to investigate whether they have more difficulty with class-maintaining affixes or with class-changing affixes.

Some examples of derivational morphological errors detected in the YELC are given below. The examples in (1) are errors of class-changing affixes (suffixes), while the examples in (2) are those of class-maintaining affixes (prefixes). These examples show that learners can make inappropriate choices of affixes or root morphemes naturally when producing sentences (see Appendix 5 for more examples of derivational morphological errors).

(1) Errors of class-changing affixes (suffixes)

- a. If there is corporal punishment, students will be quiet at the time when teachers say 'be quiet' because they fear the *sorrowness (<sorrow) of punishment.
- b. If we use the nickname *consistly (<consistently) without name, there will many problems in internet.
- c. Also, *Iranish (<Iranian) people have very positive ideas about Korea.
- d. ... drivers are so uncomfortable and *carefuless (<careless) while he is using cellular phones on driving ...

(2) Errors of class-maintaining affixes (prefixes)

- a. Using cellular phones makes us *disconcentrate (<lose concentration?) on driving.
- b. If our society let this solution used in common and legalize this solution, we don't need any animals to get some *unproper (<improper) results.

- c. Because of their *unrational (<irrational) mind, teachers can't control them with the words.
- d. As you may know, when actors or singers do *inethical (<unethical) acts, many netizens say they are very bad.

As *WordSmith Tools* detected a considerable number of derivational morphological errors of both kinds of affixes, it seems clear that these errors are among those commonly made by L2 learners. Table 9.1 shows the results of the corpus-based analysis, which indicate that Korean learners of English have more difficulties with class-maintaining affixes than with class-changing affixes. Out of 28,197 occurrences of class-changing affixes, there were 283 errors (10.0%); while out of 4,302 occurrences of class-maintaining affixes, learners had made 199 errors (46.3%). According to the G test, the error frequency of class-maintaining affixes is significantly larger than that of class-changing affixes ($G\ square = 253.92299$, $p\ value < .001$).

Table 9.1 Error frequency according to different kinds of affixes

	Class-changing affixes	Class-maintaining affixes
Total frequency	28,197	4,302
Error frequency	283	199
Error permillage (%)	10.0‰	46.3‰
Number of texts in which errors occur	253	166
Text occurrence rate (%)	89.4%	83.4%

In addition, the text occurrence rates for both types of affixes are quite high (89.4% and 83.4% respectively), which indicates that the frequency data for derivational morphological errors are not skewed by a smaller number of idiosyncratic writers.

The findings clearly show that L2 learners are likely to have more difficulty with the selection of prefixes that change meaning, rather than with the selection of suffixes that change class (part of speech), as class-maintaining affixes (prefixes) always alter the meaning of the given word, while class-changing affixes (suffixes) alter the part of speech.

Schmitt & Zimmerman (2002) suggested that L2 learners have partial knowledge of derivatives and that they appear to have more difficulty with certain parts of speech than with others. In line with this, one possible reason for the finding above might be related to the cues of part of speech in the root morphemes to which the affixes are attached. For example, the class-changing affixes (suffixes) that form adjectives, such as *-able/-ible*, *-ful*, always appear with verbs or nouns. Likewise, suffixes that form verbs, such as *-ify/-fy*, *-ize/-ise*, are always attached to nouns or adjectives. On the one hand, L2 learners seemed to be able to recognise whether the association of a particular part of speech with a particular suffix is inappropriate. This could help them to avoid derivational morphological errors. On the other hand, the class-maintaining affixes (prefixes) provide the L2 learners with more options from which to choose than do class-changing affixes. For example, the class-maintaining affix *un-* can be attached to almost every kind of content word, such as nouns, verbs, adjectives and adverbs. In that respect, the class-maintaining affixes can be regarded as much more versatile in forming derivative words. Hence, L2 learners make significantly more errors with class-maintaining affixes than with class-changing affixes, as shown in Table 9.1.

Another possible reason is that L2 learners might be affected by some specific spellings of affixes, albeit that this effect might be limited. For example, Korean learners made fewer errors with *im-*, *il-*, *ir-* than with the other class-maintaining affixes. It might be that, as the prefix *il-* is always associated with words beginning with the letter ‘*l*’, for example, L2 learners who recognise these cues would be less likely to make such errors.

However, a review of the previous studies on derivational morphological errors reveals that, to date, there is no consensus as to which kind of affixes L2 learners find most difficult. For example, Hayashi & Murphy (2011) compared the morphological awareness between class-maintaining and class-changing affixes. In the first experiment, they investigated ‘relational knowledge’ of derivational morphology, using data from a word segmentation task (i.e. data from a receptive language skill). The experiment showed that both English native speakers and Japanese learners of English performed significantly better with class-maintaining affixes than with class-changing affixes, which is in contrast to the findings of the current study. On the other hand, the second experiment dealt with productive language data obtained from an affix elicitation task that looked at ‘syntactic knowledge’ of derivational morphology. For both types of affixes the scores of Japanese learners of English did not differ statistically from those of native speakers. The differences in the results of the two experiments might be explained by the fact that not only did they investigate different aspects of morphological awareness (see Section 5.1.2 for more information about three aspects of knowledge of derivational morphology), but also, each dealt with a different language skill (receptive or productive).

Similarly, Schmitt & Meara (1997) showed that although Japanese learners of English had poor knowledge of derivative suffixes, they performed better at a task measuring receptive knowledge than at a task measuring productive knowledge. This suggests that there is a

difference in the level of cognitive processing required by the receptive task compared with the productive task (Carlisle, 2000). In other words, the productive task seems to make a greater cognitive demand on learners, than the receptive task does.

Therefore, as the findings in this section are in conflict with some of the previous studies, they seem to shed new light on the L2 acquisition of derivational morphology. The discrepancy in results may be due to characteristics of the data (receptive vs. productive) as well as the measurement tools (tests) that were used to gain the data. More specifically, the current study investigated naturally occurring data from a learner corpus (the YELC), while most previous studies employed L2 learners' data obtained from tests of receptive language skills.

9.2 The Results According to Proficiency Level

In order to identify the effect of distributional knowledge of derivational morphology according to proficiency level, this study focused on the error frequencies of three proficiency groups: low, intermediate and advanced (see Section 6.2.1). Table 9.2 shows the error per millage according to proficiency level.

As shown in Table 9.2 and Figure 9.1, for both kinds of affixes the error occurrence rate of derivational morphology gradually reduces as the proficiency level increases. The overall error per millage falls from 18.0‰ at low proficiency level, to 14.6‰ at intermediate level and finally to 6.7‰ at advanced level. For both class-changing and class-maintaining affixes, the error per millage shows similar trends to the overall error per millage.

Table 9.2 Error occurrence per millage according to proficiency level

	Low	Intermediate	Advanced	Total
Overall	116 / 6447 (18.0‰)	354 / 24252 (14.6‰)	12 / 1800 (6.7‰)	482 / 32499 (14.8‰)
Class-changing	61 / 5548 (11.0‰)	215 / 21080 (10.2‰)	7 / 1569 (4.5‰)	283 / 28197 (10.0‰)
Class-maintaining	55 / 899 (61.2‰)	139 / 3172 (43.8‰)	5 / 231 (21.6‰)	199 / 4302 (46.3‰)
Note: the figures in the cells indicate: error frequency / total frequency (error per millage)				

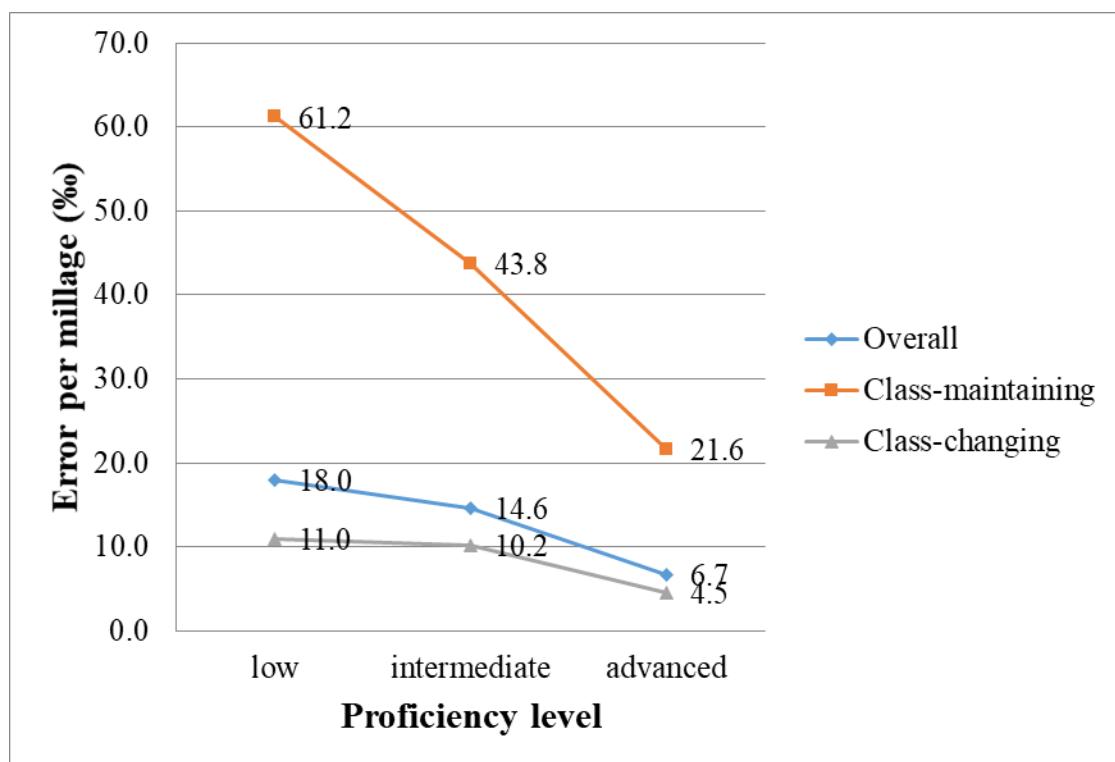


Figure 9.1 Error occurrence rate according to proficiency level

Based on the results of chi-square tests as presented in Table 9.3, there are statistically significant differences of overall error frequency between intermediate and advanced (*chi-square* = 7.04568, *p value* < .01) and between low and advanced (*chi-square* = 11.08397, *p value* < .001) levels. Although the difference between low and intermediate is not statistically significant (*chi-square* = 3.67450), it seems to be very close to the significance level (.05).

Table 9.3 Results of chi-square test comparing proficiency level groups

	Overall	Class-changing	Class-maintaining
Low vs. Intermediate	3.67450	0.19908	4.27593*
Intermediate vs. Advanced	7.04568**	4.37993*	2.09434
Low vs. Advanced	11.08397***	4.84833*	4.95367*
Note: * <i>p value</i> < .05 ** <i>p value</i> < .01 *** <i>p value</i> < .001			

With regard to class-changing affixes, the difference in error frequency between intermediate and advanced level (*chi-square* = 4.37993, *p value* < .05) and between low and advanced level (*chi-square* = 4.84833, *p value* < .05) is statistically significant. However, the difference between low and intermediate level is not significant. For class-maintaining affixes the difference in error frequency between low and intermediate (*chi-square* = 4.27593, *p value* < .05) and between low and advanced (*chi-square* = 4.95367, *p value* < .05) is statistically significant, while the difference between intermediate and advanced is not.

In summary, the differences between low and advanced levels in all three categories appear to be significant. Therefore, it could be argued that, in general, as L2 learners' proficiency in

writing improves, the error frequency falls significantly. The findings clearly show that the distributional knowledge aspect of derivational morphology awareness can be regarded as a highly relevant indicator for L2 learners' writing proficiency. Furthermore, it seems that the derivational morphological errors made in the course of L2 acquisition do not become fossilised, as the error occurrence rate significantly reduces as the proficiency level increases. In other words, the results suggest that a facilitative effect of morphological awareness (Schmitt & Zimmerman, 2002), as mentioned in Section 5.1.3, is valid not only in receptive language skills but also in productive language skills.

The results of the chi-square tests according to the writing proficiency of L2 learners in the current research echo those of Leontjev et al. (2016), whose study showed the correlation between word derivational knowledge and writing proficiency of L2 adolescent students (see Section 5.1.3). This tendency appears to be valid with regard to not only the class-maintaining affixes, but also the class-changing affixes.

On the other hand, the findings of this study are, to some extent, in contrast to those of McCutchen & Stull (2015). As mentioned in Section 5.1.3, they used the term 'morphological inventions' to refer to neologisms produced by native English speaking children. In their study, native speakers of primary school age (10 to 11 years) made more morphological inventions as their morphological awareness improved, whereas in the current study, the L2 university students made fewer errors as their proficiency level increased. This might be because the native speaking children were still developing their cognitive capability, whereas the cognition of the L2 university students (19 to 20 years old) in the current study seems to be already fully developed. Consequently, it is likely that adult L2 learners will recognise derivational morphological errors more easily than do native speaking children, as their English proficiency improves. If a similar kind of study to that of McCutchen & Stull

(2015) were to be conducted with older native speaking students, it could be predicted that the findings would be similar to the results of the current research; however, that is beyond the scope of this thesis.

9.3 Derivational Morphological Error Types According to the Degree of Overgeneralisation

As mentioned in Section 9.1, *WordSmith Tools* detected 482 derivational morphological errors in total. Detailed investigation regarding their possible causes suggests that derivational morphological errors can be categorised by the degree to which L2 learners overgeneralise the affixation rules when they choose affixes in association with the corresponding root morphemes. The degree of overgeneralisation involved in selecting specific affixes and their corresponding root morphemes cannot easily be measured. However, I propose to divide derivational morphological errors into four identifiable levels, which can be described along a cline of the degree of overgeneralisation.

Table 9.4 Categorisation of derivational morphological errors based on the corpus-based error analysis

Category	Description
Type A	Learners mistakenly select an incorrect affix, being confused with a similar form, e.g. <i>-er/-or</i> , <i>-able/-ible</i> .
Type B	Learners select an incorrect affix with a correct root morpheme.
Type C	Learners select an incorrect form of the root morpheme with a correct affix.
Type D	Learners select both an incorrect root morpheme and an incorrect affix.

Table 9.4 describes how each error type is categorised, where Type A shows a low degree, Type B (and Type C) a moderate degree and Type D a high degree of overgeneralisation. The underlying assumption of this categorisation is that the degree of overgeneralisation is closely related to the extent of involvement of ‘cognitive vehicles’ (see Section 3.4.2). In other words, it could be speculated that, in the case of Type A errors, cognitive vehicles have very little involvement, because these errors seem to stem from L2 learners’ somewhat superficial confusion of forms, whereas it appears that the cognitive vehicles are highly involved in the case of Type D errors, to an extent that the result seems strange but also very creative.

Here are some examples of Type A errors:

(3) Type A errors

- a. It is matter of how *instructer (<instructor) motivates children in doing their activities.
- b. The object that our family have specail meaning is a *scuplter (<‘sculptor’, possible mistake for ‘sculpture’).
- c. ... but the mental damage he got was *irreversable (<irreversible).
- d. So communication with others by using phone is *inevitible (<inevitable).

([sic], from the YELC)

First, it appears that Type A errors are produced when learners are confused by orthographically and phonologically similar forms, as in (3), where the confusing pairs are –*er*–*or* and –*able*–*ible*. Type A errors might derive from a simple confusion rather than a systematic process as a result of learners’ overgeneralisation of affixation rules. These errors seem to be very similar to a type of ‘synforms’ (similar lexical forms) (Laufer, 1988). As briefly mentioned in Section 3.5.2, Laufer (1988) used ‘synforms’ as an umbrella term that

encompasses ‘synphones’, ‘syngraphs’ and ‘synmorphs’, indicating that L2 learners’ errors are related to similar sounds, scripts (spellings) and morphological structures, respectively. Although the concept of synforms provides a valuable framework to categorise these kinds of errors, regrettably Laufer did not discuss the cognitive processes involved with these phenomena. In the current study, Type A covers only those errors that involve the misuse of two pairs, but it seems both novel and significant to attempt to measure the degree of overgeneralisation in a case where cognitive vehicles play only a very limited role.

(4) Type B errors

- a. It’s very *unjustice (<injustice) in victim’s case.
- b. The most series problem is *undirect (<indirect) smoking.
- c. Some people may dislike unthinkable *obeyness (<obedience).
- d. If all teachers explain student’s *disorrect (<incorrect) point and punishment.
- e. ... drivers could be a *murder (<murderer).

([sic], from the YELC)

Secondly, Type B errors are produced when learners choose incorrect affixes rather than root morphemes, as shown in (4). According to the data employed by this study, learners made many more Type B errors with class-maintaining affixes than with class-changing affixes (see Table 9.5 in this section), probably because among the class-maintaining affixes, the choice among many ‘negation’ affixes is a particular cause of confusion, as in (4a), (4b) and (4d). On the other hand, they made relatively few errors of class-changing affixes, as in (4c) and (4e). In particular, (4e) is unique, because this error involves the omission of the suffix, rather than the addition of an inappropriate suffix. It could be that the learner assumed the existence of a verb *murd*, which is converted to an agentive noun by adding the suffix *-er*.

This might look like a ridiculous mistake, but it seems to have been produced in a logical manner, in terms of the operation of the L2 learner's cognitive vehicles.

(5) Type C errors

- a. It's easy to *legisters (<legislators) because they have to only one thing – change the law.
- b. But *paradoxly (<paradoxically), I feel very cool and happy.
- c. Our Korean men must wake up and complete military service *voluteerly (<voluntarily).

([sic], from the YELC)

Thirdly, in contrast to Type B, Type C errors can be identified as those produced when L2 learners select incorrect root morphemes rather than affixes, as in (5). For example, in (5b), the suffix *-ly* should have been added to the adjective *paradoxical* rather than the noun form *paradox*. Likewise in (5c), the noun form *volunteer* is mistakenly associated with the suffix *-ly*. It could be speculated that Type B and C errors are produced according to a moderate degree of overgeneralisation.

(6) Type D errors

- a. I can't explain *detaiely (< in detail).
- b. For nation's *saveness (<safety) and peace, it's a regular to give military service.
- c. If that kind of thing is *rehappen (<happen again), there must be serious damages.
- d. ... other people think physical punishment in school is *irregal (<illegal) and is just old fashion.
- e. [many people] ... also regard smoker as a patients like *drunker (<a person who got drunk) who have a alchoholic.

f. ... people who are suffering from some kinds of *irrestorable disease (< disease that is not restorable?) can be dead because of lack of new medicines.

([sic], from the YELC)

Fourthly, Type D errors consist of a mix of Types B and C, where learners choose both incorrect affixes and inappropriate root morphemes and therefore unintentionally create new words that do not exist in English. An important factor to distinguish Type D from Type B or C errors can be the existence, or not, of corresponding target words. The target words for Type D errors do not exist in English, hence the learner needs a new phrase to indicate the presumed intended meaning, whereas Type B or C errors apparently correspond to target words that do exist. For example, although there is no such word in English as *rehappen* as in (6c), we can guess the meaning the learner intended to express. Example (6d) seems more creative. The learner might have been aware that the prefix *ir-* can be added to words beginning with ‘r’ to form a negative complex word, but took the incorrect form *regal* for the adjective *legal*. He or she followed the affixation rule but eventually made the bizarre word *irregal*. Example (6e) is also significant, in that the learner might have supposed that a person who gets drunk might be called a *drunker*, with the agentive suffix *-er* attached to the adjective *drunk*. Note that there is no such word as *drunker* according to the Collins Online English Dictionary, but the word form *drunker* occurs 11 times in the BNC, as the comparative form of *drunk*. In other words, in the same way that native speakers create new forms that did not previously exist in English, L2 learners might perform similar mental processes in a logical way. In this respect, the working of cognitive vehicles seems to be maximised in the production of Type D errors (see Appendix 5 for more examples of Type A, B, C and D).

Table 9.5 and Figure 9.2 show the frequencies and occurrence rates of derivational morphological errors according to the error type. For class-changing affixes, the error occurrence rate of Type D (4.3‰) is the highest among the four types, although it is still quite low in comparison with that of the same type for class-maintaining affixes (16.7‰). When it comes to class-maintaining affixes, Type B ranks the highest at 25.3‰, followed by Type D at 16.7‰. This indicates that L2 learners tend to make many more Type B errors with class-maintaining affixes than with class-changing affixes. Note that, according to the data used in this study, learners made a considerable number of Type D errors with class-changing as well as class-maintaining affixes.

Table 9.5 Error frequency according to error type

	Type A	Type B	Type C	Type D	Sum
Class-changing	32 (1.1‰)	41 (1.5‰)	89 (3.2‰)	121 (4.3‰)	283 (10.0‰)
Class-maintaining	0 (0.0‰)	109 (25.3‰)	18 (4.2‰)	72 (16.7‰)	199 (46.3‰)
Total	32	140	107	193	482
Note: The denominator when calculating the error per millage in the round brackets is the total token frequency of each affix (i.e. class-changing=28,197, class-maintaining=4,302)					

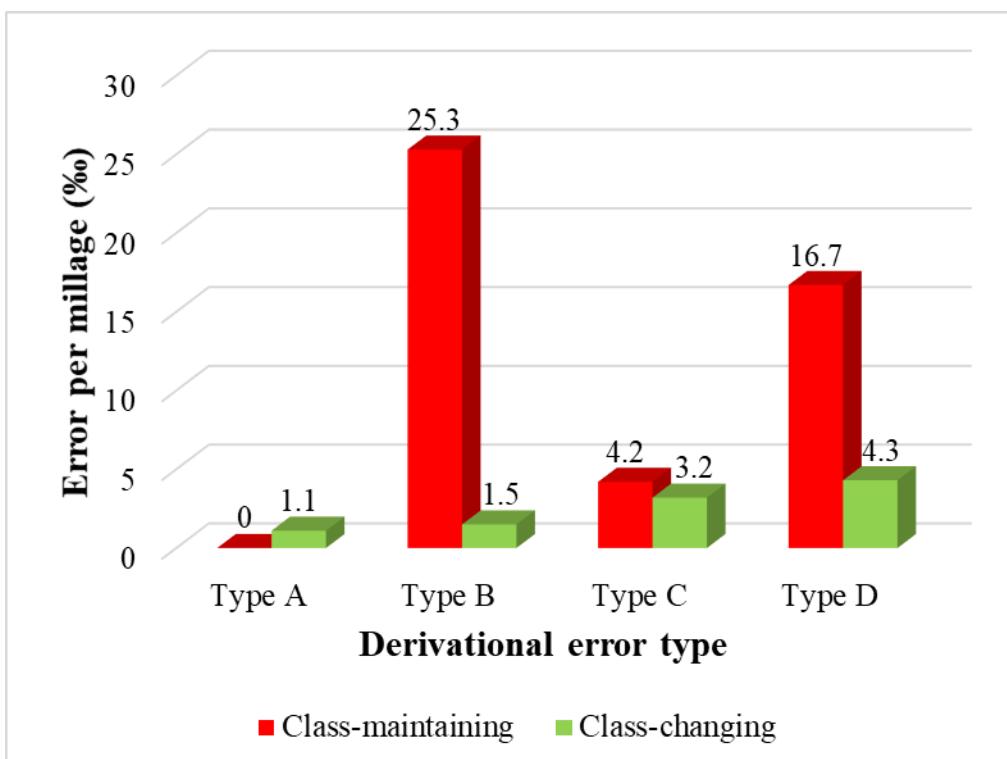


Figure 9.2 Error occurrence rates of each error type

Table 9.5 and Figure 9.2 outline how the four error types are distributed according to the type of affixes. The results are interesting, because with the one exception of Type B with class-maintaining affixes, the error occurrence rates for both kinds of affixes appear to increase from Type A to Type D. Considering that this categorisation of error types is related to the extent of involvement of ‘cognitive vehicles’ (see Table 9.4 in this section), it could be argued that Korean learners are likely to make more derivational morphological errors as they rely more on cognitive vehicles. On the other hand, the reason why the error occurrence rate of Type B with class-maintaining affixes is exceptionally high (25.3%) might derive from the characteristics of class-maintaining affixes (prefixes) investigated in this study (see Table 6.8 in Section 6.3.3). Seven class-maintaining affixes out of eight involve adding information of ‘negation’ to the complex words. This could be not only because prefixes that add negation

seem to be dominant in English (Bauer & Nation, 1993), but also because this study selected the prefixes that occur frequently in the YELC (see Section 6.3.3). In other words, L2 learners might choose incorrect prefixes because they are confused by the range of available options when it comes to prefixes that add negation.

The results presented here should not be taken as pointing to any one firm conclusion. The error categorisation used in this study is not set according to a critical theoretical background, albeit that it is still a useful and valuable criterion to identify intralingual influence on derivational morphological errors. In short, Table 9.5 and Figure 9.2 serve to give a brief overview of the error occurrence rates of class-maintaining and class-changing affixes according to the degree of overgeneralisation involved in producing the four error types.

9.4 Intralingual Influence: Overgeneralisation of Affixation Rules

According to the results of this research, the four types of derivational morphological errors, from Type A to Type D, show a gradually increasing degree of overgeneralisation of the affixation rules that apply to the L2 source. Therefore, it could be argued that there is intralingual influence on derivational morphological errors and that the errors are produced by the interplay of the L2 source and cognitive vehicles (see Section 3.5.1).

Type A and B errors derive from the incorrect choice of affix, despite the correct choice of root morpheme associated with the affix. Although these two types look similar, they differ in that Type A results from orthographic confusion, whereas Type B errors are due to incorrect selection rather than the confusion of similar forms. In contrast to Type B, Type C involves the incorrect choice of root morphemes rather than affixes. As mentioned in Section 6.3.3, simple spelling mistakes of the root morphemes are excluded from these types of errors.

As shown in the examples in Section 9.3, Types B and C seem more complicated and systematic than Type A, in that L2 learners are more likely to rely on cognitive processes that actually lead them away from the correct choice of affix and/or root morpheme. Thus, it could be argued that Type B and C errors show some evidence of the operation of cognitive vehicles, where L2 learners try to generalise the affixation rules they have learnt in the L2 acquisition.

The highest degree of L2 learners' use of cognitive vehicles is found in Type D errors, where the overgeneralisation of affixation rules leads to the coining of new words that do not exist in English. Strange though they may look, it seems that these errors are produced in a logical way, according to the underlying cognitive processes.

Interestingly, some Type D errors might be caused by interlingual influence (L1 transfer). These cases can be found with the help of the Sejong Corpus (see Sections 6.2.4 and 6.3.3). For example, the Sejong Corpus contains an equivalent word in Korean for *irrestorable* as in (6f) in Section 9.3, although there is no corresponding word in English. This suggests that the writer literally translated the Korean word into English, and he or she used the affixation rule in the course of coining the new word. The possible interlingual influence behind these errors is analysed in the next section.

9.5 Interlingual Influence Shown in Type D Errors

As noted above, the findings suggest that interlingual influence plays a role in affixation errors made by Korean learners. This is conspicuous particularly in the case of Type D errors, where learners create words that do not exist in English. This does not necessarily mean that there is no interlingual influence on errors that belong to the other types. Rather, it would be

difficult to identify the interlingual influence in, for example, Type B or C errors, because it would be very hard to claim that an incorrect selection of certain affixes or root morphemes in producing the derivational morphological errors is affected only by the L1 source. However, since the forms created by Type D errors do not exist in English, then if Korean equivalents of the incorrect forms do exist it seems reasonable to conclude that L2 learners are resorting to the L1 source when they create the errors. In other words, if the newly coined English words (errors) have Korean equivalents, then these would be regarded as errors caused by interlingual influence (L1 transfer), because it is highly likely that these words have been literally translated from Korean.

In order to identify the interlingual influence (L1 transfer) on the Type D derivational morphological errors, I made a close examination of the concordance lines of those errors. In doing so I drew mainly upon the Sejong Corpus, as mentioned in Section 6.3.3. I also referred to the Standard Korean Language Dictionary, published by the National Institute of the Korean Language. Examples of the interlingual errors of derivational morphology extracted from the YELC are shown in (7) below:

(7) Interlingual errors (from Type D)

- a. So as the education tool of their formation to socially *communicatable people, physical punishment is needed ...
- b. ... so today teacher learn from the *licenseless teacher also use physical punishment.
- c. However, this made students to behave *unrightly.
- d. ... they will die in prison unless some laws are *recorrected.
- e. My last grade of highschool vacation is very *disworth.

([sic], from the YELC)

To reiterate, the target forms shown in (7) do not exist in English. They are all newly coined by Korean learners as a result of derivational morphological errors. Interestingly, the incorrect forms have Korean equivalents and appear to be literal translations from the Korean language. Hence, it seems highly likely that the L1 is the source of these errors. For example, in (7b), by ‘licenseless teacher’ the writer seems to refer to a teacher who is not certificated by the government. Interestingly, according to the Standard Korean Language Dictionary there is a Korean word that describes the condition of having ‘no license’ (*무자격*). This could be strong evidence of literal translation. Similarly, although the adverb ‘unrightly’ as in (7c) does not exist in English, the Korean language does contain an adverb that indicates exactly that meaning (*불공정하게*). Likewise, other examples in (7) have their Korean translations according to the Sejong Corpus or the Standard Korean Language Dictionary. The evidence of literal translations shown in these examples supports the argument that there is interlingual influence on the derivational morphological errors (see Appendix 5 for more examples of interlingual influence).

Table 9.6 shows the frequency of the Type D errors, along with the frequency of the interlingual errors according to proficiency level. Out of 193 Type D errors, 40 are interlingual errors, representing 20.7% of the total. Given that the interlingual error percentage declines as proficiency level increases, it can be concluded that the interlingual influence is strongest at a low level of proficiency and weakest at advanced level.

Table 9.6 Error frequency of Type D and the interlingual influence

	Low	Intermediate	Advanced	Total
Type D error frequency	44	142	7	193
Interlingual influence	17	22	1	40
Interlingual influence / Type D (%)	38.6%	15.5%	14.3%	20.7%

In short, the examples of Type D errors as in (7) and the error frequencies presented in Table 9.6, are evidence of interlingual influence in the L2 acquisition of derivational morphology. Therefore, these errors appear to support the L2 lexical development model and the proposed new error taxonomy presented in Section 3.5.

9.6 Conclusion

This chapter has shown different types of derivational errors retrieved from the YELC, a Korean learner corpus of English that consists of naturally produced texts written by university students (see Section 6.2.1). It has revealed both interlingual and intralingual influences as possible sources of derivational morphological errors. It has also demonstrated the relationship between morphological awareness and writing proficiency.

First, the findings from naturally occurring data show that L2 learners seem to have more difficulties with class-maintaining affixes (prefixes) than with class-changing affixes (suffixes), which is somewhat in contrast to some previous studies. Thus, the results of the current study show how the L2 acquisition of derivational morphology could differ according to the type of language skill (receptive vs. productive).

Secondly, the findings in this chapter are significant in that they show that the derivational morphological errors of L2 English learners can be categorised by the degree of overgeneralisation of affixation rules. The overgeneralisation of rules can be regarded as intralingual influence. However, the errors also reveal that L2 learners' mother tongue can affect the development of morphological awareness. In particular, some Type D errors appear to be literal translations from the Korean language, according to the comparative analysis with the Sejong Corpus and the Standard Korean Language Dictionary (see Section 9.5). This is believed to be strong evidence of interlingual influence. Hence, it can be suggested that not only intralingual but also interlingual influence can play a role in the L2 acquisition of derivational morphology. In addition, it appears that the cognitive vehicles in the mental lexicon play important roles in the degree of overgeneralisation of the rules.

Thirdly, the findings clearly show that L2 learners' morphological awareness, particularly distributional knowledge of derivational morphology, is closely related to the level of proficiency in writing. In other words, as learners' morphological awareness develops, their proficiency in writing seems to increase (or vice versa). This tendency is seen most clearly with regard to class-maintaining affixes. In addition, it seems that the errors do not become fossilised.

In conclusion, this chapter has shown that derivational morphological errors can be categorised according to the sources of the errors; that is, intralingual and interlingual influences, as presented in the proposed new error taxonomy in Section 3.5.2. It has also demonstrated that the distributional knowledge of derivational morphology represented in productive learner data is correlated with L2 learners' writing proficiency and that the derivational morphological errors do not become fossilised.

CHAPTER 10: ORTHOGRAPHIC ERRORS

This chapter provides evidence of both interlingual and intralingual influences in the phonological/orthographic domain as defined in the proposed new error taxonomy presented in Section 3.5.2. Using the corpus-based error analysis, it explores how both influences play roles in the spelling errors produced by Korean learners of English. First, this chapter shows that some of those spelling errors can be traced to interlingual influence. Then, it provides evidence that other spelling errors are caused by intralingual influence. Therefore, the orthographic errors addressed in this chapter can be divided into two categories according to their source, as interlingual or intralingual errors.

In Section 10.1, I give an overview of the corpus-based error analysis, which revealed the error frequencies of both interlingual and intralingual spelling errors. In Section 10.2, I provide illustrative examples of interlingual spelling errors that seem to be influenced by three pairs of consonant substitutions (see Section 5.2.3). I also give examples of intralingual spelling errors associated with processes of inflection and derivation. In Sections 10.3 and 10.4, I investigate in more detail how each influence plays a role at different levels of English proficiency, in comparison with the occurrence rates of overall spelling errors in the YELC. I also address the question of whether or not spelling errors become fossilised. In Section 10.5, I discuss the error occurrence rate of each individual subcategory according to English proficiency level in order to show whether or not the findings in Sections 10.3 and 10.4 are skewed by any individual subcategory.

10.1 Overview of the Corpus-based Error Analysis on Spelling Errors

As mentioned in Section 5.2.3, with regard to interlingual spelling errors the research for this study focused on three pairs of consonant substitutions ([p]-[f], [l]-[r] and [b]-[v]), while in the investigation of intralingual spelling errors it focused on inflectional and derivational errors. Through the research procedures of the corpus-based error analysis presented in Section 6.3.4, this study obtained a number of both interlingual and intralingual spelling errors.

Table 10.1 Overview of spelling errors in the YELC retrieved through the corpus-based error analysis

	Error frequency	Error percentage ³³ (%)	Number of texts in which errors occur	Text occurrence rate (%)
All spelling errors	11,636	1.072%	N/A	N/A
Interlingual errors (a)	256	0.024%	221	86.3%
Intralingual errors (b)	1,479	0.136%	1,276	87.6%

As reported in Table 10.1, the overall spelling error frequency in the YELC is 11,636, which is about 1% of the whole corpus content. It thus appears that, generally speaking, Korean learners of English make at least one spelling error in every 100 words. The errors in the YELC include 256 interlingual and 1,479 intralingual errors, which represent 0.024% and 0.136% respectively of the corpus content. Note that these figures do not include all the spelling errors under interlingual and intralingual influence, as mentioned in Section 6.3.4. In

³³ The denominator of the given error percentage is the overall token number of the YELC, which is 1,085,918. The percentage figures have been rounded up to the third decimal.

addition, as the fourth and fifth columns show, the text occurrence rates for both interlingual and intralingual errors are very high (86.3% and 87.6%), hence it can be argued that the frequency data for spelling errors are not skewed by a small number of writers.

Table 10.1 clearly shows that there is not only intralingual influence, but also interlingual influence, even in the writings of L2 learners whose first language does not use the Roman orthographic system used for English. These findings are in line with previous studies that have reported both interlingual and intralingual spelling errors made by L2 learners (e.g. Salam, 2016; Al-Busaidi & Al-Saqqaf, 2015; El-Hibir & Al-Taha, 1992; Fashola et al., 1996).

In order to show the error occurrence rates for both interlingual and intralingual spelling errors more accurately, this study compared the error frequencies with the occurrences of correctly spelled words (see Section 6.3.4).

Table 10.2 Comparison between frequencies of spelling errors and correctly spelled words in the YELC according to the corpus-based error analysis

	Error frequency	Number of target forms	Frequency of correct words	Error occurrence rates ³⁴ (%)
Interlingual errors	256	124	26,066	0.973%
Intralingual errors	1,479	535	81,179	1.789%

As presented in Table 10.2, this study found the target forms of every error for both categories of spelling errors by a manual investigation: 124 target forms for interlingual errors and 535 target forms for intralingual errors. With these target forms, *WordSmith Tools* detected the frequencies of correctly spelled words that occur in the YELC: 26,066 correctly

³⁴ The denominator of the rates is the sum of the error frequency plus the frequency of correct words.

spelled words for the target forms of interlingual errors and 81,179 correctly spelled words for the target forms of intralingual errors, so that the error occurrence rates are 0.973% and 1.789% respectively. Considering that the error percentage of all spelling errors that occur in the YELC is only 1.072% (see Table 10.1), the findings show that Korean learners made slightly fewer than average interlingual spelling errors (0.973%), whereas they made many more than average intralingual spelling errors (1.789%). These results indicate that interlingual influence does play a role in producing spelling errors in the case of three pairs of consonant substitutions, although the power of the influence seems limited (*chi-square* = 2.28691, not significant³⁵). However, the intralingual influence on spelling errors related to inflection and derivation appears to be statistically significant (*G-square* = 306.41443, *p value* < .001). Therefore, it could be suggested that Korean learners are more likely to make intralingual spelling errors than interlingual spelling errors.

10.2 Examples of Interlingual and Intralingual Spelling Errors

This section illustrates examples of the spelling errors detected by the corpus-based error analysis. As mentioned in the previous section, the research for this study found 256 interlingual spelling errors and 1,479 intralingual spelling errors in the YELC.

With regard to the interlingual influence on spelling errors, this study focused on three consonant sounds and their corresponding letters. As there are no [f], [r] or [v] sounds in the Korean phonology, this study hypothesised that Korean learners are likely to confuse them with Korean [p], [l] and [b] sounds (see Section 5.2.3).

³⁵ SIGIL was utilised for the statistical measurements (see Section 6.3.2 for more information about SIGIL).

As hypothesised, Korean learners of English seem to have difficulty in discerning these sounds, which results in interlingual spelling errors involving the use of the corresponding letters. Table 10.3 illustrates some examples of the interlingual spelling errors. Owing to [p]-[f] confusion, some learners substituted the letter ‘f’ for ‘p’, e.g. **imfortant*, **ficture*. Conversely, some learners substituted the letter ‘p’ for ‘f’, e.g. **perpect*, **specipic*. There was also a substitution of ‘p’ for ‘ph’ ([f] sound), as in **physcal*. Similarly, regarding the [l]-[r] confusion, some learners substituted the letter ‘l’ for ‘r’, e.g. **ploblem*, **legister*. Conversely, other learners substituted the letter ‘r’ for ‘l’, e.g. **congraturation*, **arive*. The case of the spelling error **irregal* is interesting, because if the word ‘regal’ (for ‘legal’) were correct, the morphologically coined word **irregal* might also have existed as an antonym. In other words, this example shows how the phonological confusion may first cause an interlingual spelling error, which then leads to an incorrect choice of prefix (*ir-*) to produce a derivational morphological error, which belongs to the morphological domain (see Chapter 9). In addition, [b]-[v] confusion resulted in spelling errors such as **liverty* or **Batican* (see Appendix 6 for more examples of interlingual spelling errors).

Table 10.3 Examples of interlingual spelling errors

Type	Examples
[p] and [f] substitution	<i>*imfortant</i> (< <i>important</i>), <i>*perpect</i> (< <i>perfect</i>), <i>*specipic</i> (< <i>specific</i>), <i>*physcal</i> (< <i>physical</i>), <i>*ficture</i> (< <i>picture</i>)
[l] and [r] substitution	<i>*ploblem</i> (< <i>problem</i>), <i>*congraturation</i> (< <i>congratulation</i>), <i>*arive</i> (< <i>alive</i>), <i>*irregal</i> (< <i>illegal</i>), <i>*legister</i> (< <i>register</i>)
[b] and [v] substitution	<i>*liverty</i> (< <i>liberty</i>), <i>*lavoratory</i> (< <i>laboratory</i>), <i>*Batican</i> (< <i>Vatican</i>)

These spelling errors demonstrate the interlingual influence in the acquisition of L2 orthography, which is in line with the findings of previous studies (e.g. Salam, 2016; El-Hibir & Al-Taha, 1992; Fashola et al., 1996; James et al., 1993). Indeed, the interlingual spelling errors in the current study are very similar to those reported by Salam (2016) and El-Hibir & Al-Taha (1992), which showed that Arabic learners seem to be confused between ‘p’ and ‘b’ sounds and spellings. They are also very similar to the ‘unpredicted errors’ of Fashola et al. (1996), where Spanish learners were affected by Spanish phonology when producing English spelling errors (see Section 5.2.2).

Interestingly, these findings show that even in the phonological/orthographic domain in the L2 mental lexicon, interlingual influence plays a role, namely in the production of spelling errors, which is contrary to Jiang’s (2000) suggestion that there might be no interlingual influence in the L2 acquisition of phonological/orthographic features (see Section 3.3.3).

On the other hand, the cause of the intralingual errors is arguably the incorrect application of English spelling rules, which might not be related to L2 learners’ mother tongue. Table 10.4 presents some examples of inflectional or derivational spelling errors detected through the corpus-based analysis.

Table 10.4 Examples of intralingual spelling errors

Type	Examples
Inflectional spelling errors	*teachs (<teaches), *studing (<studying), *happend (<happened), *writting (<writing), *humen (<humans), *hitted (<hit), *controled (<controlled), *diarys (<diaries), etc.
Derivational spelling errors	*punishmet (<punishment), *dangerouse (<dangerous), *responsibillity (<responsibility), *physicaly (<physically), *quallity (<quality), *valueable (<valuable), etc.

Korean learners of English sometimes omitted a specific letter (e.g. **studing*, **controled*, **physicaly*, **punishmet*), inserted an unnecessary letter (e.g. **writting*, **quallity*, **valueable*), or incorrectly applied spelling rules, as in **humen*. The spelling error **humen* is presumably an incorrect plural form of *human* and might derive from the case of *man-men*, as a result of an incorrect application of the spelling rule. The spelling error **diarys* presumably results from the overgeneralisation of plural *-(e)s* (see Appendix 6 for more examples of intralingual spelling errors).

As in the case of interlingual spelling errors, the findings of this study with regard to intralingual spelling errors are in line with those of previous studies that have shown overgeneralisation of spelling rules in the L2 acquisition of orthography (e.g. M. Park, 2015; El-Hibir & Al-Taha, 1992; Moon & Kim, 2015; Fashola et al., 1996). For example, M. Park (2015) used a dictation test to show that Korean high school students made a number of inflectional spelling errors. El-Hibir & Al-Taha (1992) also reported that Saudi students made a number of morphological errors³⁶ in the plural formation of nouns and in the third person singular verb form. Interestingly, El-Hibir & Al-Taha (1992: 86) indicated that the cause of the morphological errors was a so-called “strategy of second language communication”, which can be regarded as a type of what the current study refers to as cognitive vehicles (see Section 3.4.2). This suggests that other researchers have also noted that some errors can be caused by these kinds of cognitive processes, along with L2 sources. In addition, Moon & Kim (2015) found that Korean university students made a number of spelling errors when adding inflectional morphemes like *-ing* and *-(e)d* and morphological morphemes such as *-ly* and *-ful*. They suggested that the errors they found from dictation tests could be caused by learners’ lack of inflectional or morphological knowledge of English. In the same vein, the

³⁶The category of ‘morphological error’ in El-Hibir & Al-Taha’s (1992) study can be regarded as the same as what this study terms ‘spelling errors’.

findings of the current study with regard to intralingual errors are very similar to the ‘predicted errors’ reported by Fashola et al. (1996) (see Section 5.2.2).

Therefore, the findings in this section appear to support the L2 lexical development model and the proposed new error taxonomy (see Section 3.5), which asserts that both interlingual and intralingual influences have effects in the phonological/orthographic domain, by providing empirical evidence of both kinds of spelling errors.

10.3 Interlingual Spelling Errors According to Proficiency Level

In order to find out how interlingual influence plays a role in the production of spelling errors according to proficiency level, this study investigated the frequencies of both errors and correctly spelled words. In the same way that I obtained the frequency data presented in Table 10.2 in Section 10.1, based on the target forms of each individual error, I collected the frequencies of correct words according to proficiency level.

Table 10.5 and Figure 10.1 present the frequency data of overall spelling errors and of interlingual spelling errors, as well as the results of statistical tests between them. Note that the frequency data of overall spelling errors are given as an average in each proficiency level for the comparison with interlingual spelling errors. With regard to the interlingual spelling errors, learners at a low level of proficiency made 114 errors, intermediate level learners made 138 errors, while at advanced level there were 4 errors. However, the error occurrence rate dropped consistently, from 1.998% at low level, to 0.717% at intermediate and 0.294% at advanced level. Similarly, the error occurrence rates of overall spelling errors appear to decrease as the English proficiency level rises from low to advanced. It is particularly interesting that the error occurrence rate of interlingual errors at a low level of proficiency is

significantly higher than that of overall errors (1.998% vs. 1.512%, G -square = 7.98923, p value < .01), whereas at an intermediate level, the error rate of interlingual errors is significantly lower than that of overall errors (0.294% vs. 0.642%, chi -square = 12.28965, p value < .001). Likewise at an advanced level, the rate of interlingual errors is 0.294%, which is lower than the rate of overall errors (0.642%), although the difference is not statistically significant.

Table 10.5 Comparison between interlingual spelling errors and overall spelling errors according to proficiency level³⁷

	Low	Intermediate	Advanced	Total
Token frequency	234,555	801,845	49,518	1,085,918
Overall spelling errors	3,547	7,771	318	11,636
Error occurrence rates of overall spelling errors	1.512%	0.969%	0.642%	1.071%
Frequency of correct words	5,593	19,115	1,358	26,066
Interlingual spelling errors	114	138	4	256
Error occurrence rates of interlingual spelling errors	1.998%	0.717%	0.294%	0.973%
Statistical tests between overall and interlingual errors	G -square = 7.98923, p value < .01	chi -square = 12.28965, p value < .001	chi -square = 2.03592, not significant	chi -square = 2.28691, not significant

³⁷ The percentage figures shown in some cells were calculated by the proportion of the error frequencies in the given proficiency levels. Therefore, the denominators of the given error percentages differ according to proficiency level. The figures have been rounded up to the third decimal.

These results show that the interlingual influence is strongest at a low level and seems to become weaker as English proficiency increases to intermediate and advanced levels. That is, Korean learners at a low level of proficiency are likely to make significantly more than the average number of interlingual spelling errors, while at the other levels, they make fewer errors than the average. These findings could provide valuable pedagogic implications for the language classroom. Teachers need to know that Korean learners of English, particularly those at a low level of proficiency, can make interlingual spelling errors and that at intermediate and advanced levels interlingual influence on spelling errors is still effective but relatively limited.

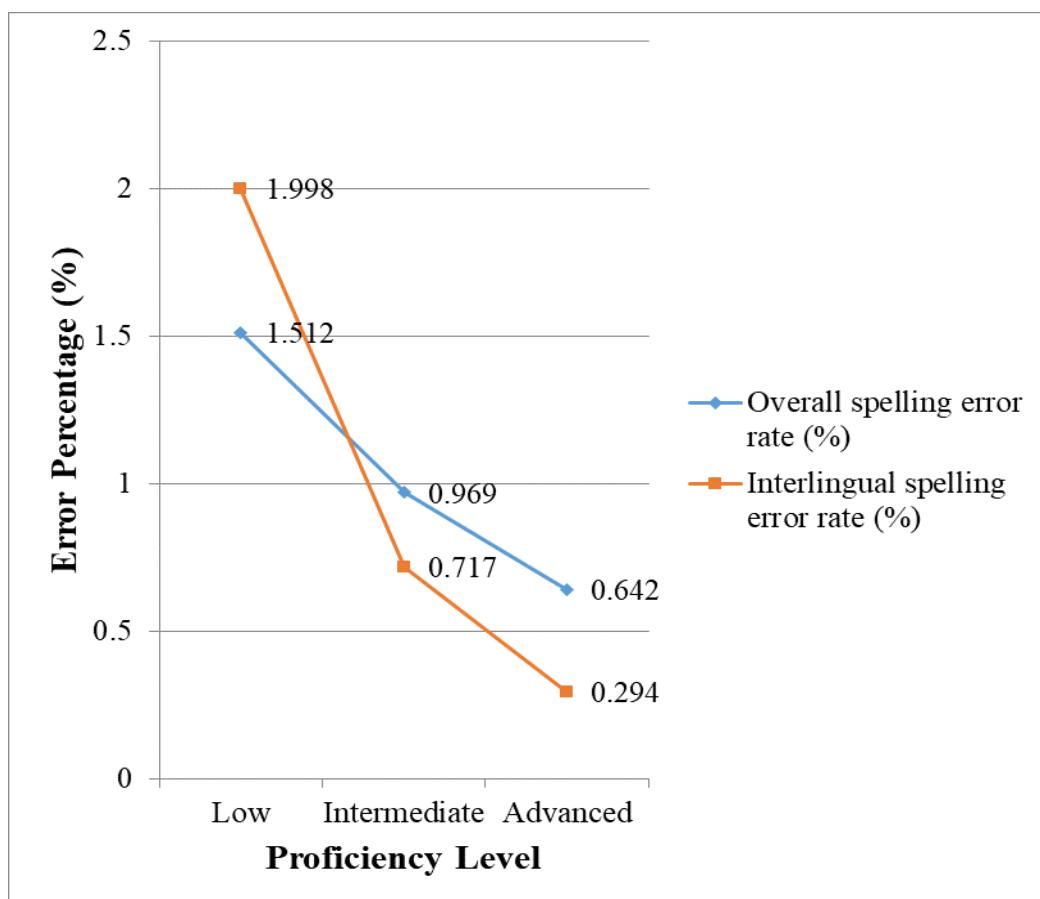


Figure 10.1 Comparison of error occurrence rates between interlingual spelling errors and overall spelling errors according to proficiency level

In addition, it is very important to find out whether or not interlingual spelling errors become fossilised as learners' English proficiency increases. This would also provide valuable pedagogic implications for teachers and learners in the English classroom in terms of how to avoid spelling errors. Table 10.6 shows the results of statistical tests for interlingual spelling errors according to the English proficiency level. The tests revealed that there are statistically significant differences between groups, except for between intermediate and advanced groups. In other words, these statistics indicate that the error frequency of a higher proficiency group is significantly lower than that of a lower proficiency group. Hence, it seems that interlingual influence on spelling errors does not become fossilised; rather, the orthographic awareness of Korean learners of English significantly improves as their English proficiency increases.

Table 10.6 Results of statistical tests for interlingual influence comparing proficiency level groups

	Interlingual influence
Low / Intermediate	<i>G-square = 61.75839, p value < .001</i>
Intermediate / Advanced	<i>chi-square = 2.73873, not significant</i>
Low / Advanced	<i>chi-square = 18.42411, p value < .001</i>

These results are not surprising, as it is to be expected that L2 learners at a low level of proficiency would make more spelling errors than those at a higher level. This might explain why researchers have rarely investigated the relationship between spelling error occurrence rates and English proficiency level, and why teaching spellings in the L2 classroom has been neglected (Westwood, 2014). However, according to the results in this section, it is worth

noting that there is interlingual influence on spelling errors, particularly at a low proficiency level.

10.4 Intralingual Spelling Errors According to Proficiency Level

As mentioned in Section 10.1, according to the results of the analysis in this study, when taking into account all spelling errors, intralingual influence is statistically significant. Therefore, this section explores in more detail how intralingual influence plays a role in spelling errors according to English proficiency level. As with the interlingual influence explored in the previous section, this study examined not only the intralingual error frequency but also the frequency of corresponding correct words, which were obtained from the target forms of intralingual spelling errors according to proficiency level. It then compared the error occurrence rates of intralingual errors at each proficiency level with those of overall spelling errors.

Table 10.7 and Figure 10.2 show the frequencies of intralingual spelling errors and correct words found in the YELC according to the English proficiency level, along with the frequencies of overall spelling errors. Table 10.7 also indicates the results of statistical tests between intralingual and overall spelling errors. The findings for intralingual spelling errors show that learners with low proficiency made 440 errors, intermediate level learners made 999 and there were 40 errors at advanced level. Similar to the case of interlingual spelling errors discussed in the previous section, the error occurrence rate decreases from 2.538% at low level, to 1.632% at intermediate and 0.972% at advanced level. The error occurrence rate of overall spelling errors also decreases as the English proficiency level rises.

Table 10.7 Comparison between intralingual spelling errors and overall spelling errors according to proficiency level³⁸

	Low	Intermediate	Advanced	Total
Error occurrence rates of overall spelling errors	1.512%	0.969%	0.642%	1.071%
Frequency of correct words	16,899	60,205	4,075	81,179
Intralingual spelling errors	440	999	40	1,479
Error occurrence rates of intralingual spelling errors	2.538%	1.632%	0.972%	1.789%
Statistical tests between overall and interlingual errors	<i>G-square = 93.49835, p value < .001</i>	<i>G-square = 213.01826, p value < .001</i>	<i>G-square = 198.43052, p value < .001</i>	<i>G-square = 306.41443, p value < .001</i>

The line graphs in Figure 10.2 clearly show that in each case, the error occurrence rate decreases as the English proficiency level of Korean learners becomes higher. However, in contrast to the interlingual influence, which is strong only at a low level of proficiency (see Table 10.5 in Section 10.3), it appears that intralingual influence remains strong at all proficiency levels. According to the statistical tests between the frequencies of overall spelling errors and intralingual errors, the differences between them at all three proficiency levels are statistically significant. This suggests that Korean learners of English are highly likely to make intralingual spelling errors, which are produced in the course of attaching inflectional or derivational morphemes to words.

³⁸ The error occurrence rates of overall spelling errors presented in the second row in Table 10.7 (and the line of the overall spelling error rate in Figure 10.2) are the same as those in Table 10.5 (and Figure 10.1). The error occurrence rates of overall spelling errors at each proficiency level in the YELC are used as averages for the comparisons with the rates of intralingual spelling errors.

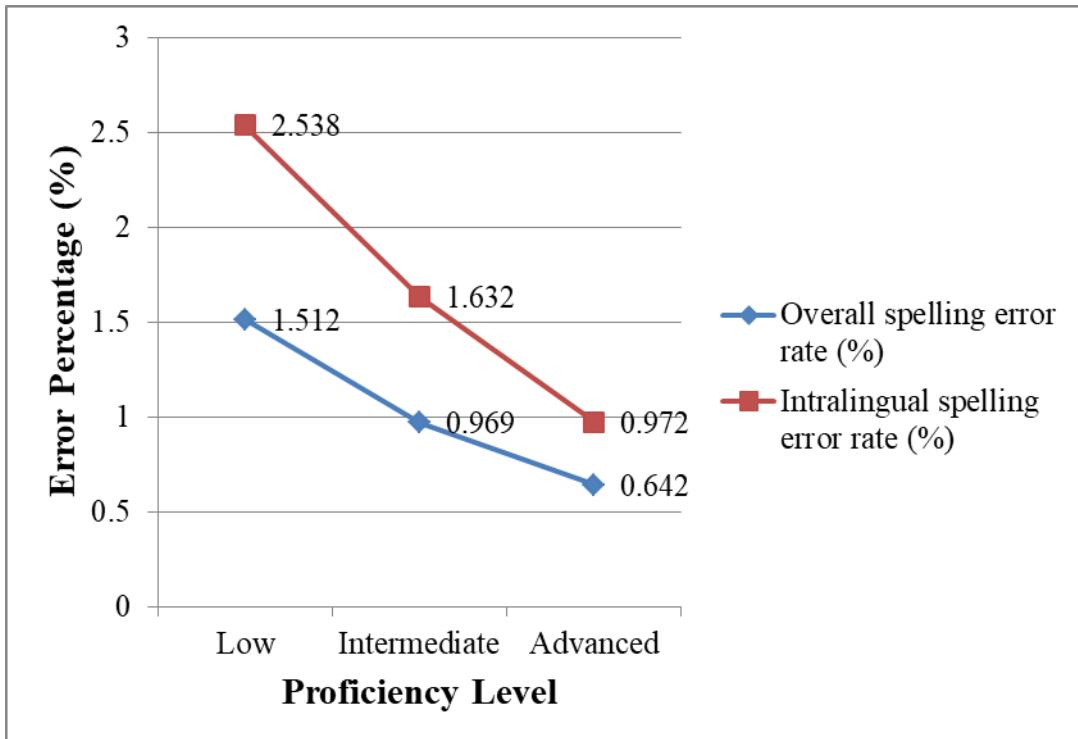


Figure 10.2 Comparison of error occurrence rates between intralingual spelling errors and overall spelling errors according to proficiency level

In addition, since the frequency of intralingual spelling errors produced by Korean learners seems to be related to English proficiency level, this study conducted a statistical analysis of the intralingual influence according to proficiency level in order to identify whether or not this influence becomes fossilised. As shown in Table 10.8, the statistical tests revealed that between each pair of two groups there are statistically significant differences. These indicate that intralingual influence on spelling errors becomes significantly weaker as learners' proficiency increases. Accordingly, it could be cautiously argued that intralingual influence does not become fossilised. However, it should also be noted that, as shown in Table 10.7 and Figure 10.2, the error occurrence rates of intralingual spelling errors produced by Korean learners are significantly higher than the overall spelling error rates at every proficiency level.

As with the findings of the previous sections and those discussed just above, these results may provide pedagogic implications for the English classroom.

Table 10.8 Results of statistical tests for interlingual influence comparing proficiency level groups

	Intralingual influence
Low / Intermediate	$G\text{-square} = 56.99863, p \text{ value} < .001$
Intermediate / Advanced	$\chi^2 = 10.31831, p \text{ value} < .01$
Low / Advanced	$\chi^2 = 36.55510, p \text{ value} < .001$

10.5 Individual Influence of Each Subcategory

This section addresses the error occurrence rates of each subcategory within both interlingual and intralingual spelling errors in order to determine whether or not the results in Sections 10.3 and 10.4 were skewed by any individual category.

Regarding the interlingual errors, Table 10.9 shows the error occurrence rates of three pairs of consonant substitutions, along with the frequencies of errors and correct words, based on the target forms of each individual error. As shown in Table 10.9 and Figure 10.3, the error rates of [p]-[f] and [l]-[r] substitutions decrease as the English proficiency level rises from low to advanced. In contrast to these two substitutions, interestingly, the error rate of [b]-[v] substitution gradually increases. This suggests that the error occurrence rate of [b]-[v] substitution skews that of interlingual spelling errors presented in Section 10.3. However, due to the relatively low frequency of the [b]-[v] substitution (see Table 10.9), it does not seem to affect the tendency of the error rate of interlingual spelling errors. The reason why Korean

learners have difficulty with [b]-[v] substitution even at an intermediate or an advanced level seems to be related to loan words from foreign languages. I discovered that among 19 target forms of interlingual spelling errors with [b]-[v] substitution, there are seven loan words, e.g. *vaccine*, *bacteria*, *veteran*, *Vatican*. Although these loan words may sometimes have equivalents in the Korean language, which sound similar to the words in English, it is likely that Korean learners might be used to seeing them in written form in Korean script (*hanguel*) rather than in English.

Table 10.9 Error occurrence rates of subcategories of interlingual spelling errors according to proficiency level

	Low	Intermediate	Advanced	Total	Number of target forms
[p] and [f] substitution	57 / 3,310 (1.693%)	63 / 11,186 (0.560%)	2 / 825 (0.242%)	122 / 15,321 (0.790%)	50
[l] and [r] substitution	50 / 2,056 (2.374%)	53 / 7,078 (0.743%)	0 / 454 (0.000%)	103 / 9,588 (1.063%)	56
[b] and [v] substitution	7 / 609 (1.136%)	22 / 1,857 (1.171%)	2 / 138 (1.429%)	31 / 2,604 (1.176%)	19

Note: the figures in the bold cells indicate: error frequency / frequency of correct words
(error occurrence rate (%))

Table 10.9 and Figure 10.3 also show that the error occurrence rates of [p]-[f] and [l]-[r] substitutions at a low level of proficiency (1.693% and 2.374% respectively) are higher than the rate of overall spelling errors (1.512%, see Table 10.5 in Section 10.3), of which results are in line with those demonstrated in Section 10.3. The rates of both substitutions drop

sharply at intermediate and advanced level, which is similar to the case of the overall interlingual spelling error rate, as presented in Figure 10.1 in Section 10.3, suggesting that the error occurrence rates of [p]-[f] and [l]-[r] substitutions do not skew the trend of the overall interlingual spelling errors.

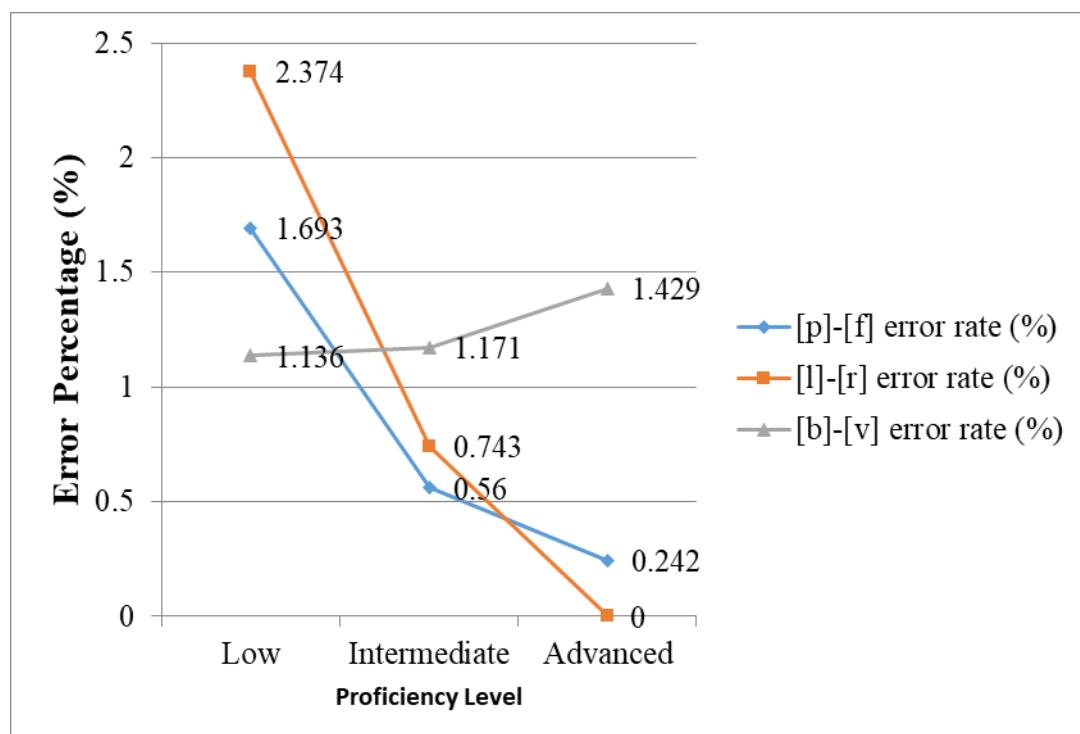


Figure 10.3 The interlingual error occurrence rates according to proficiency level

The findings shown in Table 10.9 and Figure 10.3 indicate that interlingual influence on spelling errors, particularly with [p]-[f] and [l]-[r] substitutions, decreases as the English proficiency level becomes higher. On the other hand, the influence of [b]-[v] substitution appears to increase slightly with proficiency level, which might be related to the fact that Korean learners have difficulty with loan words. These findings could contribute to raising

awareness of teachers and learners, and help to develop more sophisticated teaching materials for L2 learners.

With regard to intralingual influence, Table 10.10 and Figure 10.4 show the intralingual spelling error occurrence rate in each subcategory according to proficiency level. As can be seen from the table and figure, Korean learners of English made more intralingual spelling errors in the process of inflection than in derivation, at all three proficiency levels.

Table 10.10 Error occurrence rates of subcategories of intralingual spelling errors according to proficiency level

	Low	Intermediate	Advanced	Total	Number of target forms
Inflectional spelling errors	296 / 8920 (3.212%)	633 / 33770 (1.840%)	23 / 2252 (1.011%)	952 / 44942 (2.074%)	306
Derivational spelling errors	144 / 8367 (1.692%)	366 / 28186 (1.282%)	17 / 1970 (0.856%)	527 / 38523 (1.350%)	242
Note: the figures in the bold cells indicate: error frequency / frequency of correct words (error occurrence rate (%))					

Furthermore, it can be argued that the error frequency data of the intralingual spelling errors presented in Section 10.4 are not skewed by any individual subcategory. This is supported by the fact that, as shown in Figure 10.4, the error occurrence rates of both inflectional and derivational spelling errors decrease as proficiency level rises from low to advanced, which indicates a similar trend to the rate of the intralingual spelling errors (see Figure 10.2 in Section 10.4).

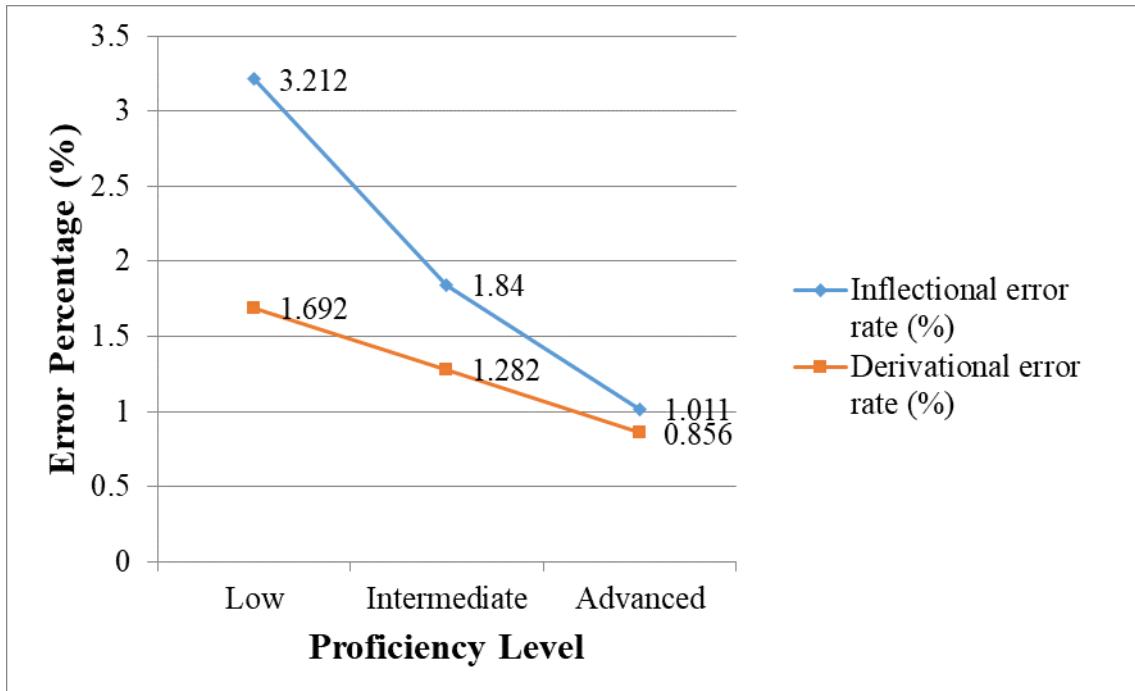


Figure 10.4 The intralingual error occurrence rates according to proficiency level

In summary, the investigation of error occurrence rate of each subcategory of both interlingual and intralingual spelling errors according to proficiency level has found that, generally speaking, there is no individual influence on spelling errors. Although there is one exception, that of [b]-[v] substitution, the effect appears to be limited because the error frequency of [b]-[v] substitution is relatively low (see Table 10.9). However, it is worth noting that the error rate of [b]-[v] substitution shows an opposite trend, presumably because Korean learners have difficulty with loan words.

10.6 Conclusion

This chapter has provided evidence of both interlingual and intralingual influences in the phonological/orthographic domain by investigating both types of spelling errors in the YELC: three pairs of consonant substitution errors for L1 phonological interference

(interlingual influence); inflectional and derivational spelling errors for intralingual influence (see Section 5.2.3).

Firstly, this study investigated interlingual spelling errors with three pairs of consonant substitutions through the corpus-based error analysis. The findings clearly show that there can be L1 phonological influence in the L2 acquisition of orthography. This research is significant in that it shows interlingual influence using evidence from naturally occurring data written by L2 learners whose mother tongue does not use the Roman orthographic system, but uses its own script, *hanguel*. This is the reason why the current research focused on the L1 phonological influence rather than the L1 orthographic influence (see Section 5.2.2). Furthermore, the interlingual influence with [p]-[f] and [l]-[r] substitutions appears to be quite strong at a low proficiency level and becomes weaker as the English proficiency level rises to intermediate and advanced. However, the effect with [b]-[v] substitution shows the opposite trend, which might derive from the difficulty with loan words.

Secondly, the findings with regard to intralingual spelling errors add to and confirm those of previous research by showing that L2 learners tend to make spelling errors in the process of inflection and derivation. The intralingual influence seems to be stronger than the interlingual influence. It also appears to be consistent at all three proficiency levels. Considering that the intralingual spelling errors are derived from the L2 source and cognitive processes; that is, they are the result of incorrect application of orthographic rules in English, the findings could provide valuable insights for the language classroom for L2 learners whose L1 does not use the Roman orthographic system.

In conclusion, this chapter has shown that spelling errors produced by L2 learners can be categorised into two different kinds, namely interlingual or intralingual spelling errors,

according to their source, as hypothesised in the proposed new error taxonomy in Section 3.5.2. In addition, this chapter has shown that both interlingual and intralingual spelling errors do not become fossilised, but reduce as English proficiency level increases. These findings could provide valuable pedagogic implications for the language classroom toward the development of relevant teaching materials.

CHAPTER 11: CONCLUSION

11.1 Main Findings of the Study

This study has addressed two main research questions: 1) How do interlingual and intralingual influences affect the production of L2 learners' lexical errors? 2) Is it possible to categorise the lexical errors according to their sources and domains presented in the proposed new error taxonomy? (See Sections 1.3 and 6.1)

To answer the research questions, the study has presented the L2 lexical development model and proposed a new error taxonomy based on the model (see Section 3.5). With regard to RQ1, it selected and investigated one error feature from each lexical domain in the taxonomy in order to show both interlingual and intralingual influences on lexical errors. With regard to RQ2, the findings of the study provide relevant empirical evidence and clearly support the grounds for the proposed L2 lexical development model and the new error taxonomy (see Table 11.1).

First, in the semantic domain, this study investigated collocational errors of three pairs of dimensional adjectives: *large-small*, *high-low* and *long-short*. It was found that Korean learners of English make deviant expressions such as *large accident* or *small attention* (see Section 7.2). Corpus-based error analysis, using reference corpora as control corpora, suggested that these errors are affected by interlingual influence and could be regarded as evidence of 'semantically incongruent' areas between L1 and L2 in the L2 mental lexicon. Similarly, it was found that Korean learners produce collocational errors such as *small confidence* or *high effect* (see Section 7.4) and that these seem to be affected by intralingual influence. In addition, this study has shown that interlingual influence can sometimes help L2

learners to acquire collocations which may be semantically congruent between L1 and L2 (positive L1 transfer) (see Section 7.3).

Table 11.1 The new error taxonomy according to lexical domains and sources (adapted from Table 3.1 in Section 3.5.2)

Source Domain	Interlingual error (L1 source)	Intralingual error (L2 source)
Semantic domain (collocational errors with dimensional adjectives)	Literal translation from collocations in Korean (e.g. <i>large accident, small attention</i>)	Approximate selection within the limited range of L2 vocabulary (e.g. <i>small confidence, high effect</i>)
Syntactic domain (over-passivisation errors of non-alternating unaccusative verbs)	L1 morphological influence (i.e. significant differences in the numbers of errors between ‘matched’ and ‘mismatched’ verbs)	Subject animacy effect (i.e. more errors with animate subjects than with inanimate subjects)
Morphological domain (derivational morphological errors)	Literal translation with Type D errors (e.g. <i>communicatable, licenseless, unrightly</i>)	Overgeneralisation of affixation rules (e.g. four types according to the degree of overgeneralisation)
Phon/Orth domain (spelling errors)	L1 phonological influence of three pairs of consonant substitution (e.g. <i>imfportant, ploblem, congraturation</i>)	Incorrect application of orthographic rules with inflectional and derivational morphemes (e.g. <i>studing, hitted, valueable</i>)

Secondly, in the syntactic domain, this study examined L1 morphological influence (interlingual influence) and the subject animacy effect (intralingual influence). Statistical measurements of the incidence of over-passivisation errors of non-alternating unaccusative verbs revealed significant differences in the numbers of errors between three morphologically ‘matched’ verbs: *exist, happen* and *arrive*, and three ‘mismatched’ verbs: *occur, remain* and

appear. This showed that there is L1 morphological influence in over-passivisation errors of non-alternating unaccusative verbs (see Section 8.2). In a similar manner, the study found that Korean learners of English make more passivisation errors with animate subjects than with inanimate subjects, although this result was not statistically significant (see Section 8.5). Furthermore, the findings indicate that over-passivisation errors of non-alternating unaccusative verbs do not become fossilised, as the error frequencies decline with increase in proficiency level (see Section 8.3).

Thirdly, in the morphological domain, this study investigated the L2 acquisition of distributional knowledge of derivational morphology. Focusing upon eight class-changing and seven class-maintaining affixes, it located derivational morphological errors in the YELC (see Section 9.1). It then divided the errors into four types according to the degree of overgeneralisation of affixation rules (intralingual influence) (see Section 9.3). The findings revealed particular errors in the Type D category, which seem to be created by Korean learners through literal translation (see Section 9.5) and thus suggest that interlingual influence plays a role in producing derivational morphological errors. In addition, it was found that morphological awareness with respect to distributional knowledge of derivational morphology is correlated with writing proficiency and that these errors do not seem to become fossilised (see Section 9.2).

Fourthly, in the phonological/orthographic domain this study focused on spelling errors, looking particularly at the interlingual influence of L1 phonology and the intralingual influence in the process of inflection and derivation. It was found that Korean learners of English make interlingual spelling errors in producing three pairs of consonant substitutions ([p]-[f], [l]-[r] and [b]-[v]), which seem to be related to the L1 phonological influence. The study also found that Korean learners make intralingual spelling errors when using

inflectional and derivational morphemes, which might derive from the incorrect application of orthographic rules for English (see Sections 10.1 and 10.2). In addition, the findings indicate that both interlingual and intralingual spelling errors do not become fossilised (see Sections 10.3 and 10.4).

To summarise, with regard to RQs 1 and 2, the study has shown both interlingual and intralingual influences involved in the production of four error features, one from each lexical domain. The empirical evidence illustrated in Chapters 7 to 10 through the corpus-based error analysis clearly supports the grounds for the proposed L2 lexical development model and the new error taxonomy.

11.2 Pedagogic Implications for English Language Teaching

The findings of this study provide valuable pedagogic implications for English language teaching (ELT). Just as traditional EA studies investigated learners' interlanguage in order to gain insights as to which features of a target language pose a greater learning burden (Corder, 1981), LCR such as the current study also contributes to ELT, through discovering the errors made by L2 learners (Meunier, 2002). Therefore, I would argue that, as a basic principle, L2 learners should be taught in ways that raise awareness of errors and their possible causes. The findings of this study have shown that L2 learners' errors are caused by interlingual and intralingual influences (see Chapters 7 to 10). This might be a very good starting point for language teachers to discuss how to raise L2 learners' awareness. Then, teachers can apply exercises based on a contrastive analysis of the two languages. In the subsections that follow, I offer specific pedagogic recommendations for each feature based on the findings.

11.2.1 Pedagogic Implications for Collocational Errors of Dimensional Adjectives

Chapter 7 of this thesis illustrated collocational errors in the use of dimensional adjectives in the YELC. Through a corpus-based method, errors were detected in the use of three pairs of dimensional adjectives: *large-small*, *high-low* and *long-short*. Then, the study revealed the significant sources of dimensional adjective errors from a perspective taken from models and ideas from cognitive linguistics. By comparing Korean learners' collocations of dimensional adjectives with English and Korean reference corpora, this study presented two (or three) possible sources in the use of dimensional adjectives: negative/positive interlingual influence and intralingual influence. It is worth noting that interlingual influence is not always an obstacle for L2 learners. As discussed in Section 7.3, Korean learners can also benefit from positive interlingual influence, particularly when the dimensional adjective collocations are semantically congruent between L1 and L2. Bahns (1993) argued that, owing to the benefits of adopting a contrastive approach to lexical collocations, students have to concentrate only on the collocations that are semantically incongruent, not on those that are semantically congruent. However, the results from Chapter 7 show that L2 learners make both interlingual and intralingual errors. Therefore, L2 learners should focus not only on semantically incongruent collocations, but also on congruent ones, because they might not know which are which at the time of learning. Consequently, the textbook designers should be aware that it would be very useful for students to be provided with exercises that contain a mix of correct and incorrect collocations, so that they might identify which are correct and why. These kinds of activities would effectively raise the students' awareness of how the L1 can influence their use of the L2 in both a negative and a positive manner.

11.2.2 Pedagogic Implications for Non-alternating Unaccusative Verb Errors

As noted in Chapter 8, this study has shown that interlingual influence (L1 morphology) plays a statistically significant role in over-passivisation errors in the use of non-alternating unaccusative verbs. Since L1 morphological influence is extensive in the initial stage when learning non-alternating unaccusative verbs, language teachers should be aware of this situation and prepared to give remedial instruction. Hwang (1999) suggests that in certain conditions explicit learning is more effective than implicit learning for the L2 acquisition of English unaccusative verbs. However, the current teaching environments in Korea do not seem to be promising, especially with respect to teaching materials. It has been reported that there is no explicit instruction on or explanation of unaccusative verbs in English textbooks (Chung, 2011). This situation should be addressed so that students are given relevant instructions and explanations about unaccusative constructions in teaching materials.

11.2.3 Pedagogic Implications for Derivational Morphological Errors

As noted in Chapter 9, this study has shown that L2 learners' cognitive vehicles (see Section 3.4.2) can help them to increase their vocabulary by means of applying affixation rules. In contrast, these cognitive vehicles can also be an obstacle, if they lead to students overgeneralising the rules. It is therefore very important for language teachers to know the causes of errors, so that they not only encourage students to explore new vocabulary, but also monitor the learning process and give students guidelines on affixation rules.

It is also worth noting that, according to the findings reported in Chapter 9, productive knowledge of derivational morphology involves higher cognitive demand than does receptive knowledge. Therefore, teachers should ensure that students' acquisition of morphological

awareness reaches the level of productive knowledge, particularly in the initial stage of language learning in which they make a number of derivational morphological errors. Students should be provided with as many opportunities as possible to practise and produce words by adding prefixes or suffixes. For example, in order to raise awareness of derivational morphology, textbook designers should be aware that L2 learners tend to overgeneralise affixation rules and to be affected by interlingual influence. Designers could provide some tips by illustrating derivational errors that are frequently made by L2 learners, or by presenting some explanations of this phenomenon.

11.2.4 Pedagogic Implications for Orthographic Errors

As noted in Chapter 10, this study has shown that orthographic errors can be categorised into two different types according to the source: intralingual or interlingual spelling errors. Since the source of intralingual spelling errors is not learners' L1 but the target language (English), it can be inferred that these errors could be made not only by native English speakers but also by L2 learners whose L1 is not Korean. This suggests that intralingual spelling errors result from the incorrect application or the absence of English spelling rules (Westwood, 2014), although it does not necessarily mean that language learners should learn a set of English spelling rules in order to remedy the errors. Because many rules are very complicated and may apply to only a small number of words, and there are many exceptions to each rule, it would not be very efficient to learn spelling rules in the classroom. Instead, with regard to the efficient teaching of spelling, researchers must pay more attention to learning strategies such as frequent reminders, cognitive strategy training, invented spellings and use of mnemonics (see Gentry, 1982; Schlagal, 2002).

On the other hand, the source of the interlingual spelling errors is the Korean language. As noted in Chapter 10, this study has focused on three kinds of spelling errors with specific consonants, which clearly show the interlingual influence of L1 phonological mediation on English orthography (see Section 5.2.2). Since the interlingual spelling errors are unique, it is very important for Korean learners of English to realise that they are liable to make these kinds of errors as a result of interlingual influence.

According to Westwood (2014), there have been two main problems with the ‘traditional approach’ to teaching spelling in the classroom over the last three decades. Firstly, under the traditional approach, spelling should never be taught through explicit instruction. Instead, students in the classroom have been guided to learn to spell merely by engaging in writing to communicate their ideas. As a result, secondly, students have resorted to rote memorisation without recognising that there are more effective and systematic ways to learn how to spell.

Recently, researchers have highlighted the use of cognitive strategies, which include methods focused on the brain, such as devising, selecting, differentiating and organising patterns of letters in words, and then monitoring the strategies (metacognition). The role of the teacher is to establish effective teaching materials and to encourage students to use these kinds of strategies (Reed, 2012; Morris & Smith, 2011; Joshi, Treiman, Carreker & Moats, 2009). Recent studies have shown that the use of cognitive strategies and explicit instructions has significant effects on spelling ability or orthographic knowledge (see McNeill, 2018; Purvis, McNeill & Everatt, 2016). The use of such cognitive strategies is in line with the cognitive vehicles in the current study. It is interesting that cognitive vehicles can also be a cause of spelling errors. Therefore, it seems very important for language teachers to guide students to use them in positive ways.

In summary, this section has aimed to provide valuable pedagogic implications for the English classroom, based on the study findings. It is recommended that future teaching materials should be designed to include remedial instructions, which reflect the corpus evidence to inform the teaching of the error-prone features. The exercises can be aimed at raising awareness of error types and their possible causes. Language teachers should be aware of both interlingual and intralingual influences and the errors associated with them, so that they can provide more efficient and valuable English classes for students.

11.3 Strengths of the Study

Among the contributions made by this study, I would like to point out three strengths in particular:

- 1) The combination of learner corpus data and corpus-based error analysis has a significant impact on EA.
- 2) The approach influenced by models and ideas taken from cognitive linguistics to learners' errors provides a new perspective for EA.
- 3) The corpus-based method applied in this study can bring benefits to research in related linguistic fields.

The following sections will address each of these points in more detail.

11.3.1 The Value of Authentic, Naturally Occurring Data and of Corpus-based Error

Analysis

This study has used learner corpus data in the form of essays written by Korean learners of English. Such data, the direct product of L2 learners' output, can be considered as more authentic than receptive data obtained from, for example, grammaticality judgement tasks produced in predetermined test settings (see Sections 4.2.2 and 5.1.2). The combination of a large set of authentic and productive learner data and the corpus-based method sheds new light on EA, bringing new resources and an innovative method to research on L2 acquisition (Granger, 2002, 2009; Granger et al., 2015).

First of all, because this approach is based on a very large dataset, it could be able to provide objective, reliable and so generalisable findings. In contrast, while the findings of previous EA studies have contributed to the understanding of L2 acquisition, in the majority of those studies researchers used relatively small datasets to find and classify specific types of errors. For example, with regard to the semantic domain, the current study has used reference corpora to determine collocational errors of dimensional adjectives that occur in the YELC (see Section 6.3.1), whereas in the majority of previous studies investigating semantic errors researchers detected errors in small datasets through their intuition as native English speakers, which might be a somewhat subjective or unreliable way to determine such errors (e.g. Dušková, 1969; James, 1998; Laufer, 1988; Llach, 2011; Richards, 1974; Zimmermann, 1987; Ander & Yıldırım, 2010).

Secondly, learner corpus data that result from productive skills of interlanguage can be very useful resources in the research of L2 acquisition, especially given that in the current situation experimental tasks used to gain receptive data of L2 learners remain dominant.

Productive data are significant because for L2 learners the cognitive burden to ‘produce’ language seems to be greater than that to ‘comprehend’. Hence such data enable us to look at the same phenomena from a different perspective (Leech, 1998; Granger et al., 2015). For example, in the investigation of over-passivisation errors of non-alternating unaccusative verbs, the findings of the current study on intralingual influence (subject animacy effect) differ from those of previous studies (e.g. Chung, 2014; Pae et al., 2014), which obtained data from grammaticality judgement tests (see Section 8.5). Similarly, with regard to the investigation of the L2 acquisition of derivational morphology, although the aspect of distributional knowledge is considered a key element to measure morphological awareness (Kuo & Anderson, 2006), only a few previous studies have dealt with productive data, probably because it was difficult to gain such relevant data. However, the current study has analysed distributional knowledge, and the findings from a new perspective could make a significant contribution to the understanding of L2 acquisition of derivational morphology (see Chapter 9).

In short, the current study has brought a new viewpoint to the L2 acquisition research, by providing four corpus-based error analyses with a large dataset of L2 learners’ productive interlanguage.

11.3.2 Cognitive Linguistic Perspective on L2 Learners’ Errors

This study has identified the sources of given errors, which is the first step to providing relevant feedback to L2 learners in the ELT context. To do so, it has presented the L2 lexical development model and the new error taxonomy (see Sections 3.5.2), which speculate as to the sources of errors and cognitive processes (in this study termed ‘cognitive vehicles’, see

Section 3.4.2). Although some previous EA studies have attempted to classify lexical errors with so-called diagnosis-based criteria, namely, the factors of the errors (e.g. Dušková, 1969; Richards, 1974; Ringbom, 1987, 2007; Zimmermann, 1987), these studies have been subject to limitations such as very small datasets and fuzzy error categorisation (see Section 2.1.1). It would seem that, in contrast to the current study, previous researchers have tended to lack clarity and consistency in their approaches to the cognitive processes in L2 learners' mental lexicon (see Section 2.1.5). In particular, previous studies have sometimes treated the sources of errors and mental processes as interchangeable or without clear definition, which might result in somewhat fuzzy categorisations. The current study has attempted to provide clearer error categories by separating the sources of errors and the activity of cognitive vehicles (see Section 3.4.2).

In addition, the current study has expanded the notion of 'incongruency' between L1 and L2 beyond the semantic domain to the syntactic, morphological and phonological/orthographic domains, by demonstrating interlingual influence (L1 transfer) in these domains (see Section 2.1.1). Although it has been acknowledged that language transfer could occur not only in semantics but also in syntax and even in phonetics and phonology (Odlin, 1989), most previous studies have focused mainly on L1 semantic influence, classifying errors under such categories as 'literal translation' or 'calque' (James, 1998; Llach, 2011). The results of the current study would seem to indicate that the process of L2 acquisition could be quite different from that of L1 acquisition and that interlingual influence may affect the L2 acquisition in different lexical domains.

11.3.3 Implications of Corpus-based Error Analysis for Other Research

The corpus-based error analysis methods applied in this study may have practical implications for research in related fields.

First, because the YELC rates each essay based on the refined version of the CEFR (see Section 6.2.1), this study has been able to investigate whether the error features (except collocations of dimensional adjectives) become fossilised. Comparison of error occurrence rates according to proficiency level revealed that the given error features do not appear to become fossilised. The research method used here might not be a strict version of the longitudinal study that shows developmental patterns of given features, in that this study did not collect the error frequencies produced by the same L2 learners in certain periods. However, it can be regarded as a type of “quasi-longitudinal study” because it has dealt with “data from a homogeneous group of learners at different levels of proficiency” (Granger, 2002: 11). Consequently, this study has significance in the field of LCR, where to date there have been only a few longitudinal or quasi-longitudinal studies (e.g. Granger, 1999; Thewissen, 2013; Dagneaux et al., 1998).

Secondly, the corpus-based method utilised to identify so-called deviant expressions related to collocations of dimensional adjectives in the semantic domain can be applied in similar kinds of investigations regarding deviance in other types of representation affected by semantic incongruity (see Section 2.2.1). As noted in Sections 4.1.3 and 4.1.4, the deviant expressions or collocational errors identified in this study could be regarded as ‘figurative overextensions’ or as unique but acceptable expressions from an ELF perspective. Because English reference corpora arguably offer an objective means to compare the degree of deviance and the use of an L1 reference corpus can enable us to identify L1 influence that

may affect the given deviant expressions, the method used in this study can be adopted to conduct a wide variety of comparative analyses (see Littlemore, 2009). For example, it could be employed in a contrastive analysis of metaphor or metonymy between different languages.

11.4 Limitations of the Study

While this research makes a number of valuable contributions, as presented in the previous section, it also has some limitations.

First, the representativeness of corpora is always a potential issue in LCR. While the English reference corpora (the BoE and the BNC, see Sections 6.2.2 and 6.2.3 respectively) used in this study might not be criticised, because the BoE is one of the largest general corpora and the BNC is a ‘balanced’ general corpus, the Korean reference corpus (the Sejong Corpus, see Section 6.2.4) and the learner corpus (the YELC, see Section 6.2.1) are regrettably somewhat smaller than might be expected. The size of the Sejong Corpus is 37 million words, which could be considered small for a reference corpus. Although the YELC (1 million words) is quite large compared to other learner corpora available,³⁹ corpus-based research generally needs a large dataset, particularly with regard to the investigation of collocation (Walker, 2008). In the research of collocational errors of dimensional adjectives, this study found only 129 error occurrences out of 1,640 tokens of six dimensional adjectives (see Section 7.1). It was necessary to discard other dimensional adjectives, such as *narrow* and *shallow*, from the analysis because of the low occurrences of those adjectives in the YELC. Similarly, in the research of derivational morphological errors, this study was unable to select certain affixes because they do not occur with sufficiently high frequency in the YELC (see Section 6.3.3).

³⁹ For example, the ICLE, a pioneering collection of learner corpora, consists of sub-corpora that contain only up to 200,000 words from various L1 backgrounds (Granger, 2003).

If this study had been able to use a learner corpus larger than the YELC, the findings would have been more robust. This weakness has to be considered when interpreting the conclusions of the research. However, it should also be noted that the Sejong Corpus and the YELC are the largest corpora currently available for LCR in the Korean context.

Secondly, this study has considered only four error features – one from each of the four lexical domains – as evidence to discuss the grounds for the L2 lexical development model and the proposed error taxonomy. It is possible that one may not be able to find both interlingual and intralingual influences in other error features, not because the two influences do not play roles in those error features, but because those roles are not overt. For example, the researcher found very few error features in the syntactic domain that show interlingual influence, probably because there are very few overt similarities in the surface structure of English and Korean. On the other hand, in the investigation of spelling errors, the very large number of similarities in vowel sounds and the many differences in orthographic representations of the sounds between the two languages were obstacles to the contrastive research. If more relevant error features in the lexical domains are examined, this will contribute to confirming the grounds for the L2 lexical development model and the proposed new error taxonomy.

Thirdly, although learners' 'errors' should be distinguished from 'mistakes', which are regarded as slips or lapses in performance, it is difficult or even almost impossible to find errors differentiated from mistakes in learner corpus data, because of the nature of such data (see Section 2.1.6). In addition, the current study has a limitation with regard to error detection. Although it has consulted reference corpora as a control when detecting deviant expressions, this procedure could have been more reliable if human raters had been used to

detect the errors, especially for semantic errors. These concerns should be taken into account when interpreting the findings of this study.

Finally, I have argued that this study has particular significance because it has analysed productive data of learners' interlanguage. It seems that language research with productive data has been somewhat neglected and so should be promoted in certain areas in order to have a balanced perspective on L2 acquisition (Leech, 1998). However, the caveat is that research with productive data is not always plausible or feasible, or even desirable. For example, with regard to over-passivisation errors, this study investigated only non-alternating unaccusative verbs, because in the case of alternating unaccusative verbs whose passivised forms can be grammatically correct, it would be necessary to ask learners their intentions in order to clarify the grammaticality of every single instance, which would be extremely difficult if not impossible with learner corpus data (see Section 4.2.3). Therefore, although the significance of productive data should not be underestimated, it is also necessary to consider the limitations of such data.

11.5 Suggestions for Future Research

Based on the main findings, the strengths and the limitations of the current study, the following suggestions and recommendations for future research can be made:

First, to overcome the limitations mentioned in the previous section, future studies could extend the scope of research to many other error features, to investigate both interlingual and intralingual influences in the four lexical domains. For example, researchers could investigate other kinds of expressions with certain parts of speech, or figurative expressions such as metaphor or metonymy that may be involved with semantic incongruity between L1 and L2

(e.g. Littlemore, 2009, 2010). One might also investigate specific syntactic constructions that are represented differently in L1 and L2, such as delexical verb (or light verb) constructions (e.g. Liu, 2010; Nesselhauf, 2005; Wang, 2016) or psych verb constructions (e.g. Guilloteaux, 2001; Hartshorne et al., 2015; White et al., 1999). In the morphological domain, instead of derivational morphology one might be able to examine both interlingual and intralingual influences in the representations of other components of morphological awareness such as inflectional or compounding morphology (see Section 5.1.1). In the phonological/orthographic domain, one might investigate, for example, those influences on spelling errors involved in specific vowel sounds rather than consonants (see Section 5.2.3). In addition, the error detecting procedure could be made more reliable through the use of human raters along with reference corpora. Both inter-rater and intra-rater reliability tests could be employed in the methodology for the corpus-based error analysis.

Secondly, it would be worthwhile to implement the same sort of research with data of learners whose L1 is not Korean. For example, Chinese or Japanese learners of English might be expected to show different kinds of interlingual errors, although they probably produce similar kinds of intralingual errors. Such research would help to broaden the understanding of L2 acquisition among learners who have different mother tongues.

Finally, one of the ultimate goals of EA studies is to give relevant feedback to learners, so that they can improve their language proficiency. The findings of this study offer valuable information for designers of teaching materials to create more efficient and appropriate teaching resources. It is hoped that the findings will be reflected in innovative teaching materials. Indeed, there is a “considerable mismatch between naturally occurring English and the English that is put forward as a model in pedagogical descriptions” (Römer, 2006: 126).

Overall, the findings of this study are valuable for learners to recognise the errors, for language teachers to give relevant feedback and instructions to students and for textbook and curriculum designers to create more sophisticated and innovative teaching materials.

References

- Abutalebi, J., Cappa, S. F., & Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 497-515). Oxford: Oxford University Press.
- Adams, M. J. (1994). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Ahn, J. H. (2015). L1 Korean L2 English learners' overpassivization errors with unaccusative verbs. *Korean Journal of Linguistics*, 40(4), 633-660.
- Aitchison, J. (2012). *Words in the mind*. Oxford: Blackwell Publishers.
- Al-Busaidi, S., & Al-Saqqaqaf, A. H. (2015). English spelling errors made by Arabic-speaking students. *English Language Teaching*, 8(7), 181-199.
- Al-Shormani, M. Q., & Al-Sohbani, Y. A. (2012). Semantic errors committed by Yemeni University learners: Classifications and sources. *International Journal of English Linguistics*, 2(6), 120-139.
- 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.
- Ander, S., & Yıldırım, Ö. (2010). Lexical errors in elementary level EFL learners' compositions. *Procedia Social and Behavioral Sciences*, 2, 5299-5303.
- Anderson, J. R. (1976). *Language, memory and thought*. Hillsdale: Lawrence Erlbaum.
- Anglin, J. M. (1993). *Vocabulary development: A morphological analysis*. Chicago: Society for Research in Child Development.
- Apel, K., & Werfel, K. (2014). Using morphological awareness instruction to improve written language skills. *Language, Speech, and Hearing Services in Schools*, 45, 251-260.
- Arab-Moghadam, N., & Sénéchal, M. (2001). Orthographic and phonological processing skills in reading and spelling in Persian/English bilinguals. *International Journal of Behavioral Development*, 25(2), 140-147.
- Bae, Han Suk. (2015). Metalinguistic awareness in vocabulary acquisition: Evidence from grade six Korean EFL learners. *Primary English Education*, 21(3), 5-22.
- Bahns, J. (1993). Lexical collocations: a contrastive view. *ELT Journal*, 47(1), 56-63.

- Balcom, P. (1997). Why is this happened? Passive morphology and unaccusativity. *Second Language Research*, 13, 1-9.
- Barcroft, J. & Sunderman, G. L. (2008). Introduction: Bringing together the how and the what of the bilingual mental lexicon. *The Mental Lexicon*, 3(3), 279-288.
- Barnbrook, G., Mason, O., & Krishnamurthy, R. (2013). *Collocation: applications and implications*. London: Palgrave Macmillan.
- Barnitz, J. G. (1982). Orthographies, bilingualism and learning to read English as a second language. *The Reading Teacher*, 35(5), 560-567.
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253-279.
- Bebout, L. (1985). An error analysis of misspellings made by learners of English as a first and as a second language. *Journal of Psycholinguistic Research*, 14(6), 569-593.
- Botley, S., & Dillar, D. (2007). Investigating spelling errors in a Malaysian learner corpus. *Malaysian Journal of ELT Research*, 3, 74-93.
- Brown, H. D. (2007). *Principles of language learning and teaching*. New York: Pearson Education.
- Burt, M., & Kiparsky, C. (1972). *The gooficon: a repair manual for English*. Rowley: Newbury House Publishers.
- Burzio, L. (1986). *Italian syntax: A government-binding approach*. Dordrecht: Reidel Publishing Company.
- Callies, M. (2015). Learner corpus methodology. In S. Granger, G. Gilquin & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 35-55). Cambridge: Cambridge University Press.
- Cameron, L. (1999). Identifying and describing metaphor in spoken discourse. In L. Cameron & G. Low (Eds.), *Researching and applying metaphor* (pp. 104-132). Cambridge: Cambridge University Press.
- Carlisle, J. (1996). An exploratory study of morphological errors in children's written stories: Impact on reading. *Reading and Writing: An Interdisciplinary Journal*, 8, 61-72.
- Carlisle, J. (2000). Awareness of the structure and meaning of morphologically complex words: Impacts on reading. *Reading and Writing: An Interdisciplinary Journal*, 12, 169-190.
- Carr, P. (1999). *English phonetics and phonology: an introduction*. Oxford: Blackwell Publishing.

- Carrió, M. L., & Mestre, E. M. (2014). Lexical errors in second language scientific writing: Some conceptual implications. *International Journal of English Studies*, 14(1), 97-108.
- Chan, A. Y. W., & Li, D. C. S. (2000). English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. *Language, Culture and Curriculum*, 13(1), 67-85.
- Channell, J. (2013). Psycholinguistic considerations in the study of L2 vocabulary acquisition. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 83-96). London: Longman.
- Chen, H.-C., & Leung, Y.-S. (1989). Patterns of lexical processing in a non-native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 316-325.
- Cho, E., & Tong, F. (2014). Morphology-based cross language transfer from Korean to literacy acquisition in English and Chinese. *Asian Journal of Humanities and Social Sciences*, 2(1), 21-36.
- Cho, InJung. (2005). English-Korean translation: Passive construction. *The Journal of Translation Studies*, 6(1), 121-142.
- Choi, Yeon Hee. (2015). Roles of L1 and L2 derivational morphological awareness in L2 reading through the mediation of L2 vocabulary knowledge. *The Journal of Asia TEFL*, 12(3), 81-114.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT press.
- Chomsky, N., & Halle, M. (1991). *The sound pattern of English*. Cambridge, London: The MIT Press.
- Chung, T. (2011). The input of L2 English unaccusative verbs: EFL textbooks and teachers' knowledge. *Korean Journal of English*, 11(2), 437-464.
- Chung, T. (2014). Multiple factors in the L2 acquisition of English unaccusative verbs. *IRAL*, 52(1), 59-87.
- Chung, T. (2015). L2 acquisition of English unaccusative verbs: A longitudinal study. *English Teaching*, 70(4), 3-26.
- Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65-110). New York: Academic Press.
- Cook, V. J. (1997). L2 users and English spelling. *Journal of Multilingual and Multicultural Development*, 18(6), 474-488.
- Corder, S. P. (1967). The significance of learner's errors. *IRAL*, 5(4), 161-170.

- Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.
- Corder, S. P. (1983). Strategies of communication. In C. Færch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 15-19). London: Longman.
- Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 308-325). Oxford: Oxford University Press.
- Croft, W., & Cruse, D. A. (2012). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Crystal, D. (2003). *English as a global language*. Cambridge: Cambridge University Press.
- Csábi, S. (2004). A cognitive linguistic view of polysemy in English and its implications for teaching. In M. Achard & S. Niemeier (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 233-256). Berlin: Mouton de Gruyter.
- Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System*, 26(2), 163-174.
- Da Rocha, F. J. (1980). On the reliability of error analysis. *Studia Anglica Posnaniensia*, 12, 83-90.
- De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13(1), 1-24.
- De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism*, 1(1), 17-32.
- De Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 191-214). Amsterdam: John Benjamins.
- De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition. *SSLA*, 19, 309-329.
- De Groot, A. (1993). Word-Type effects in bilingual processing tasks: support for a mixed representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27-51). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- De Groot, A., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology*, 53A(2), 397-428.
- Deignan, A. (1999a). Corpus-based research into metaphor. In L. Cameron & G. Low (Eds.), *Researching and applying metaphor* (pp. 199-199). Cambridge: Cambridge University Press.

- Deignan, A. (1999b). Metaphorical polysemy and paradigmatic relations: A corpus study. *Word*, 50(3), 319-338.
- Dewaele, J.-M. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics*, 19(4), 471-490.
- Dewart, M. H. (1979). The role of animate and inanimate nouns in determining sentence voice. *British Journal of Psychology*, 70, 135-141.
- Díaz-Negrillo, A. (2009). *EARS: A user's manual*. Munich: LINCOM.
- Díaz-Negrillo, A., & Thompson, P. (2013). Learner corpora: Looking towards the future. In A. Díaz-Negrillo, N. Ballier & Thompson, P. (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 9-29). Amsterdam: John Benjamins.
- Dixon, R. M. W. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton.
- Dolgormaa, U., & Lee, I. (2011). L2 acquisition of English unaccusative and unergative verbs by Mongolian learners. *Language and Linguistics*, 52, 143-168.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. New York: Oxford University Press.
- Durgunoglu, A. Y., & Roediger, H. L. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language*, 26, 377-391.
- Dušková, L. (1969). On the sources of errors in foreign language learning. *IRAL* 7, 11-35.
- Ehri, L. C. (1993). How English orthography influences phonological knowledge as children learn to read and spell. In R. J. Scholes (Ed.), *Literacy and language analysis* (pp. 21-43). Hillsdale, NJ: Lawrence Erlbaum Associates.
- El-Hibir, B. I., & Al-Taha, F. M. (1992). Orthographic errors of Saudi students learning English. *The Language Learning Journal*, 5(1), 85-87.
- Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A*, 46(3), 533-558.
- Ellis, N. C. (2008). Words and their usage: commentary on the special issue on the bilingual mental lexicon. *The Mental Lexicon*, 3(3), 375-385.
- Ellis, R. (1991). Grammaticality judgements and second language acquisition. *Studies in Second Language Acquisition*, 13, 161-186.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (1997). *Second language acquisition*. Oxford: Oxford University Press.

- Ellis, R. & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139-155.
- Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Færch, C., & Kasper, G. (1986). Cognitive dimensions of language transfer. In E. Kellerman & M. Sharwood Smith (Eds.), *Crosslinguistic influence in second language acquisition* (pp. 49-65). Oxford: Pergamon Press Ltd.
- Fashola, O. S., Drum, P. A., Mayer, R. E., & Kang, S.-J. (1996). A cognitive theory of orthographic transitioning: predictable errors in how Spanish-speaking children spell English words. *American Educational Research Journal*, 33(4), 825-843.
- Ferraro, G., Nazar, R., Alonso Ramos, M., & Wanner, L. (2014). Towards advanced collocation error correction in Spanish learner corpora. *Language Resources & Evaluation*, 48, 45-64.
- Figueredo, L. (2006). Using the known to chart the unknown: A review of first-language influence on the development of English-as-a-second-language spelling skill. *Reading and Writing*, 19, 873-905.
- Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 116-129.
- Galeote, M., & Checa, Y. E. (2005). Meaning acquisition of dimensional adjectives: The case of ‘width’ dimension. *Journal for the Study of Education and Development*, 29(4), 371-386.
- Gelman, S. A., Ravn, K. E., & Maloney, L. T. (1985). When “Big” does not refer to overall size: Dimensional adjectives in context. *Papers and Reports on Child Language Development*, 24, 62-69.
- Gentry, J. R. (1982). An analysis of developmental spelling in GNYST AT WRK. *The Reading Teacher*, 36(2), 192-200.
- George, H. V. (1972). *Common errors in language learning: insights from English*. Rowley: Newbury House Publishers.
- Goh, Gwang-Youn. (2011). Corpus-based vocabulary studies in Korea: limitations and suggestions. *The Linguistic Association of Korea Journal*, 19(1), 41-60.

- Goodwin, A., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects of literacy outcomes for school-age children. *Scientific Studies of Reading*, 17(4), 257-285.
- Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology*, 39(2), 295-319.
- Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 207-220). Amsterdam: Elsevier.
- Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a development perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 267-283). Amsterdam: John Benjamins Publishing Company.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S. (1999). Use of tenses by advanced EFL learners: evidence from an error-tagged computer corpus. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora. Studies in honour of Stig Johansson* (pp. 191-202). Amsterdam & Atlanta: Rodopi.
- Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, S. Petch-Tyson, & J. Hung (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.
- Granger, S. (2003). The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, 37(3), 538-546.
- Granger, S. (2004). Computer learner corpus research: current status and future prospects. In U. Conner & T. Upton (Eds.), *Applied corpus linguistics: a multidimensional perspective* (pp. 123-145). Amsterdam: Rodopi.
- Granger, S. (2008). Learner corpora in foreign language education. In N. Van Deusen-Scholl & N. H. Hornberger (Eds.), *Encyclopedia of language and education, 2nd edition, volume 4: second and foreign language education* (pp. 337-351). New York: Springer.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13-32). Amsterdam: John Benjamins.

- Granger, S. (2012). How to use foreign and second language learner corpora. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 7-29). Chichester: Blackwell Publishing Ltd.
- Granger, S., Gilquin, G., & Meunier, F. (2015). Introduction: learner corpus research – past, present and future. In S. Granger, G. Gilquin & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 1-5). Cambridge: Cambridge University Press.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *The International Corpus of Learner English*. Version 2. Handbook and CD-Rom.
- Green, D. W. (1993). Toward a model of L2 comprehension and production. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 249-277). Amsterdam: John Benjamins.
- Green, L., McCutchen, D., Schwiebert, C., Quinlan, T., Eva-Wood, A., & Juelis, J. (2003). Morphological development in children's writing. *Journal of Educational Psychology*, 95, 752-761.
- Guilloteaux, M. J. (2001). Psychological verbs: comprehension-based vs. production-based grammar instruction. *English Teaching*, 56(4), 93-112.
- Gut, Ulrike. (2007). First language influence and final consonant clusters in the new Englishes of Singapore and Nigeria. *World Englishes*, 26(3), 346-359.
- Ha, Myung-Jeong. (2016). Linking adverbials in first-year Korean university EFL learners' writings: a corpus-informed analysis. *Computer Assisted Language Learning*, 29(6), 1090-1101.
- Ha, S., Johnson, C. J., & Kuehn, D. P. (2009). Characteristics of Korean phonology: Review, tutorial, and case studies of Korean children speaking English. *Journal of Communication Disorders*, 42, 163-179.
- Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning*, 58(1), 1-31.
- Han, Z. (2003). Fossilisation: From simplicity to complexity. *International Journal of Bilingual Education and Bilingualism*, 6(2), 95-128.
- Han, Z. (2006). Fossilization: Can grammaticality judgement be a reliable source of evidence? In Z. Han & T. Odlin (Eds.), *Studies of fossilization in second language acquisition* (pp. 56-82). Clevedon, UK: Multilingual Matters.
- Hartshorne, J., Pogue, A., & Snedeker, J. (2015). Love is hard to understand: the relationship between transitivity and caused events in the acquisition of emotion verbs. *Journal of Child Language*, 42(3), 467-504.

- Hayashi, Y., & Murphy, V. (2011). An investigation of morphological awareness in Japanese learners of English. *Language Learning Journal*, 39(1), 105-120.
- Hemchua, S., & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect*, 21(3), 3-25.
- Heo, Yong. (2004). A contrastive study on the interference of inter-phonology. *Journal of Korean Language Education*, 15(1), 235-257.
- Hinkel, E. (2002). Why English passive is difficult to teach (or learn). In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 233-259). New Jersey: Lawrence Erlbaum Associates.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund-Prytz, Y. (2008). *Corpus linguistics with BNCweb – a practical guide*. Frankfurt: Peter Lang.
- Hong, A. L., Rahim, H. A., Hua, T. K., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *The Southeast Asian Journal of English Language Studies*, 17, 31-44.
- Hong, Y-P. (2009). The 21st Century Sejong Plan: Performances and challenges. In National Institute of Korean Language (Ed.), *The new Korean life*, 19(1) (pp. 73-86). Seoul: Nulpoom Plus Press.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.
- Hubbard, P., & Hix, D. (1988). Where vocabulary meets grammar: Verb subcategorization errors in ESL writers. *CATESOL Journal*, 1, 89-100.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hwang, Jong-Bai. (1999). L2 acquisition of English unaccusative verbs under implicit and explicit learning condition. *English Teaching*, 54(4), 145-176.
- Hwang, Jong-Bai. (2006). Why do Korean EFL learners overpassivize English unaccusative verbs? *English Language Teaching*, 18(1), 195-221.
- Hwang, Y., & Choi, J. (2016). Looking into the 21st Century Sejong Corpus: Using IT hub. In National Institute of Korean Language (Ed.), *The new Korean life*, 26(2) (pp. 5-33). Seoul: Nulpoom Plus Press.
- Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development*, 9, 67-84.
- James, C. (1998). *Errors in language learning and use: exploring error analysis*. London; New York: Longman.

- James, C., Scholfield, P., Garrett, P., & Griffiths, Y. (1993). Welsh bilinguals' English spelling: An error analysis. *Journal of Multilingual and Multicultural Development*, 14(4), 287-306.
- Jarema, G. & Libben, G. (2007). Introduction: Matters of definition and core perspectives. In G. Jarema & G. Libben (Eds.), *The mental lexicon: core perspectives* (pp. 1-6). Oxford: Elsevier.
- Jarvis, S. (2009). Lexical transfer. In A. Pavlenko (Ed.), *The bilingual mental lexicon: interdisciplinary approaches* (pp. 99-124). Bristol: Multilingual Matters.
- Jeon, Eun Hee. (2011). Contribution of morphological awareness to second-language reading comprehension. *The Modern Language Journal*, 95(2), 217-235.
- Jeong, Soon-Yong, & Cho, Mi-Hui. (2016). An acoustic analysis of Korean EFL students' English vowel production: Focused on 'new' English vowel categories. *Korean Journal of English Language and Linguistics*, 16(2), 405-424.
- Jeong, Y. M. (2013). *Error types and trends of English spelling in creative writing by Korean elementary school students*. MA thesis. Kyungpook National University.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47-77.
- Jiang, N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: selection, acquisition, and testing* (pp. 101-126). Amsterdam: John Benjamins.
- Joshi, R. M., Treiman, R., Carreker, S., & Moats, L. C. (2009). How words cast their spell: Spelling is an integral part of learning the language, not a matter of memorization. *American Educator*, 32(4), 6-43.
- Ju, M. K. (2000). Overpassivization errors by second language learners: The effect of conceptualizable agents in discourse. *Studies in Second Language Acquisition*, 22, 85-111.
- Kachru, Braj B. (1985). Standards, codification and sociolinguistics realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.
- Kan, L. (2014). *Morphological awareness and vocabulary knowledge in second-language adolescent learners: an intervention study*. M.Soc.Sc thesis, University of Hong Kong.
- Katamba, F., & Stonham, J. (2006). *Morphology*. Basingstoke, UK: Palgrave Macmillan.

- Kieffer, M. J., & Lesaux, N. K. (2012). Effects of academic language instruction on relational and syntactic aspects of morphological awareness for sixth graders from linguistically diverse backgrounds. *The Elementary School Journal*, 112(3), 519-545.
- Kim, Chin-Do. (2000). Contrastive investigation of the semantic structure of the dimensional adjectives in Korean and German. *Korean Society of German Language and Literature*, 14, 41-67.
- Kim, Eokjo. (2011). A study on the meaning through combination between dimensional adjectives and preceding elements. *International Language and Literature*, 23, 75-97.
- Kim, Eun Hee. (2010). *A prototype approach to L2 acquisition of English unaccusative verbs*. MA thesis. Korea University.
- Kim, J. E., & Yoo, I. W. (2015). A corpus-based study of to-infinitive errors in Korean college freshmen's writing. *Journal of Asia TEFL*, 12(4), 37-60.
- Kim, Jeong-ryeol, & Son, Jeong-Eun. (2012). The analysis of spelling error through dictation activity in the elementary school level. *Journal of the Korea English Education Society*, 11(1), 141-162.
- Kim, Jeong Woo. (1996). *English as Korean, Korean as English: Theory*. Seoul: Changmoonsa.
- Kim, Jin-soo, & Wu, Jin-ji (2014). A study on the difference and the reason of the difference of the Korean ‘keu-ta’ and Chinese ‘da’. *The Society for Korean Language & Literary Research*, 79, 31-52.
- Kim, W. H. (1998). The passive voice: form and function. *Korean Linguistics*, 8, 131-157.
- Kim, Woo-young. (2004). The effects of focus-on-form instructions on the acquisition of English unaccusative verbs. *Foreign Languages Education*, 11(4), 419-445.
- Koda, K. (2000). Cross-linguistic variation in L2 morphological awareness. *Applied Psycholinguistics*, 21, 297-320.
- Koh, Sungran, & Kim, Bonggyu. (2014). L2 learner corpus analysis of modal expressions in written text. *English21*, 27(1), 209-229.
- Kondo, T. (2005). Overpassivization in second language acquisition. *IRAL*, 43, 129-161.
- Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 53-81). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: the role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory*, Vol. 2 (pp. 389-395). London: Wiley.

- Kroll, J. F., & Sholl, A. (1992). Lexical and conceptual memory in fluent and nonfluent bilinguals. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 191-204). Amsterdam: Elsevier.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149-174.
- Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representations and processing: looking back and to the future. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psychological approaches* (pp. 531-553). Oxford: Oxford University Press.
- Kuo, L., & Anderson, R. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, 41(3), 161-180.
- Kwon, Heokseung. (2008). The practice and application of corpus linguistics. *Korean Journal of Applied Linguistics*, 24(3), 1-30.
- Kwon, Heokseung. (2009). The SNU Korean learner corpus of English: compilation and application. *English Language and Linguistics*, 28, 203-228.
- La, Ki-Yun. (2001). Improvement methods of teaching primary school English pronunciation in the EFL environment. *Studies in English Education*, 6(2), 5-31.
- Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202-251). Cambridge: Cambridge University Press.
- Lam, B. P., & Sheng, L. (2016). The development of morphological awareness in young bilinguals: Effects of age and L1 background. *Journal of Speech, Language, and Hearing Research*, 59, 732-744.
- Lardiere, D. (2006). Knowledge of derivational morphology in a second language idiolect. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006)* (pp. 72-79). Somerville, MA: Cascadilla Proceedings Projects.
- Laufer, B. (1988). The concept of ‘synforms’ (similar lexical forms) in vocabulary acquisition. *Language and Education* 2(2), 113-132.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 74(4), 440-448.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners’ English. *Language Learning*, 62(2), 647-672.

- Lauzon, Kory, & Song, Sanghoun. (2016). Korean EFL learners can hardly understand hardly: a corpus-based study. *Korean Journal of English Language and Linguistics*, 16(4), 745-771.
- Laws, J., & Ryder, C. (2014). Getting the measure of derivational morphology in adult speech: A corpus analysis using *MorphoQuantics*. *Language Studies Working Papers*, 6, 3-17.
- Lee, D., Yeon, J., Hwang, I., & Lee, S. (2010). KKMA: A tool for utilizing Sejong Corpus based on relational database, *Journal of KIISE: Computing Practices and Letters*, 16(11), 1046-1050.
- Lee, Eun-Joo. (2017). The development of English personal reference in the Korean ESL learner corpus. *International Information Institute*, 20(3B), 2061-2072.
- Lee, H., & Choi, I. (2011). What causes overpassivization? *The Linguistic Association of Korea Journal*, 19(3), 91-107.
- Lee, Hye-Kyung, & Kim, Hyeyon-Okh. (2017). Features of lexical bundles in academic writing by non-native vis-a-vis native speakers of English. *English Language and Linguistics*, 23(2), 67-88.
- Lee, Hyungu. (2013). An error analysis of English vowel spelling-to-sound correspondences of Korean university students: based on one syllable word. *Journal of Linguistic Studies*, 18(3), 99-120.
- Lee, J-H. (2009). L1 morphological transfer as a source of errors with unaccusative verbs?: L2 acquisition of English and Korean. *Korean Journal of Applied Linguistics*, 25(3), 275-301.
- Lee, Kyu-Jin. (1997). *A study of possible pronunciation errors made by the Korean student of English though interference: with special reference to the phonological difference between the two languages*. PhD dissertation. Konkuk University.
- Lee, Sun. (2010). A study on English spelling instruction based on the analysis of invented spellings of the Korean elementary students. *Foreign Languages Education*, 17(1), 235-256.
- Lee, Yoon. (2004). Phonemic awareness of Korean primary school students. *Foreign Languages Education*, 11(3), 103-131.
- Lee, Yoon. (2007). Aspects on English spelling of Korean elementary school students. *Foreign Languages Education*, 14(3), 231-256.
- Lee, Youngsook. (1990). A contrastive analysis of passive phrases. *The Journal of Studies in Language*, 9, 61-85.

- Leech, G. (1998). Learner corpora: what they are and what can be done with them. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xx). London: Longman.
- Lempert, H. (1990). Acquisition of passive: the role of patient animacy, salience, and lexical accessibility. *Journal of Child Language*, 17, 677-696.
- Lennon, P. (1991a). Error: some problems of definition, identification, and distinction. *Applied Linguistics* 12(2), 180-195.
- Lennon, P. (1991b). Error and the very advanced learner. *International Review of Applied Linguistics in Language Teaching*, 29(1), 31-43.
- Leontjev, D., Huhta, A., & Mäntylä, K. (2016). Word derivational knowledge and writing proficiency: How do they link? *System*, 59, 73-89.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: The MIT Press.
- Levelt, W. J. M. (1999). Producing spoken language: a blueprint of the speaker. In P. Haggy & C. M. Brown (Eds.), *The neurocognition of language* (pp. 83-122). Oxford: Oxford University Press.
- Levenston, E. A. (1979). Second language acquisition: issues and problems. *Interlanguage Studies Bulletin*, 4(2), 147-160.
- Li, S. (2015). *A corpus-based study of the high frequency nouns time and thing*. PhD dissertation. University of Birmingham.
- Li, Yi-syuan, & Chen, Shu-hui. (2016). Relative effectiveness of phonological and morphological awareness training on L2 word reading in EFL children. *System*, 60, 93-104.
- Littlemore, J. (2009). *Applying cognitive linguistics to second language learning and teaching*. London: Palgrave Macmillan.
- Littlemore, J. (2010). Metaphoric competence in the first and second language: similarities and differences. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 293-315). Amsterdam: John Benjamins Publishing Company.
- Littlemore, J., & Low, G. (2006). Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics*, 27(2), 268-294.
- Liu, Chen-Pin. (2010). EFL learners' use of lexical collocations involving the light verbs *have*, *make*, and *take*. *Hwa Kang English Journal*, 16, 91-116.
- Llach, M. P. A. (2011). *Lexical errors and accuracy in foreign language writing*. Bristol: Multilingual Matters.

- Lowie, W. (2000). Cross-linguistic influence on morphology in the bilingual mental lexicon. *Studia Linguistica*, 54(2), 175-185.
- Luelsdorff, P. A. (1986). *Constraints on error variables in grammar: Bilingual misspelling orthographies*. Amsterdam: John Benjamins Publishing Company.
- Maloney, L. T., & Gelman, S. A. (1987). Measuring the influence of context: The interpretation of dimensional adjectives. *Language and Cognitive processes*, 2(3), 205-215.
- Maratsos, M. P. (1973). Decrease in the understanding of the word “Big” in preschool children. *Child Development*, 44(4), 747-752.
- Maratsos, M. P. (1974). When is a high thing the big one? *Developmental Psychology*, 10(3), 367-375.
- Mauranen, A. (2004). Spoken corpus for an ordinary learner. In J. H. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 89-105). Amsterdam: John Benjamins Publishing Company.
- McBride-Chang, C., Cho, J.-R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., Cheuk, C. S-M., & Muse, A. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea and the United States. *Journal of Experimental Child Psychology*, 92, 140-160.
- McCutchen, D., & Stull, S. (2015). Morphological awareness and children’s writing: accuracy, error, and invention. *Read Writ*, 28, 271-289.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. Abingdon: Routledge.
- McNeill, B. (2018). Improving preservice teachers’ phonemic awareness, morphological awareness and orthographic knowledge. *Australian Journal of Teacher Education*, 43(1), 28-41.
- Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper & A.P.R. Howatt (Eds.), *Interlanguage* (pp. 225-239). Edinburgh: Edinburgh University Press.
- Meara, P., & English, F. (1987). *Lexical errors and learners’ dictionaries (Reports-Research/Technical 143)*. Retrieved from ERIC Institution of Educational Sciences website:
http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1c/4b/98.pdf.
- Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, S. Petch-Tyson, & J. Hung (Eds.), *Computer learner*

corpora, second language acquisition, and foreign language teaching (pp. 119-141). Amsterdam: John Benjamins.

Meuter, R. F. I. (2005). Language selection in bilinguals: mechanisms and processes. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 349-370). Oxford: Oxford University Press.

Mitton, R. (1996). *English spelling and the computer*. Harlow, Essex: Longman Group.

Mo, J. (2015). Individual verb differences in Chinese learners' acquisition of English non-alternating unaccusatives. *Journal of Language Teaching and Research*, 6(2), 366-372.

Montrul, S. (1999). Causative errors with unaccusative verbs in L2 Spanish. *Second Language Research*, 15(2). 191-219.

Montrul, S. (2000). Transitivity alternations in L2 acquisition: Toward a modular view of transfer. *SSLA*, 22, 229-273.

Montrul, S. (2001). First-language-constrained variability in the second-language acquisition of argument-structure-changing morphology with causative verbs. *Second Language Research*, 17, 144-194.

Moon, An-Nah, & Kim, Myungsook. (2015). An analysis of English spelling errors made by Korean adult learners. *Modern Studies in English Language & Literature*, 59(4), 73-102.

Morris, H., & Smith, S. (2011). *33 ways to help with spelling: Supporting children who struggle with basic skills*. Oxon: Routledge.

Murre, Jaap M. J. (2005). Models of monolingual and bilingual language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 154-169). Oxford: Oxford University Press.

Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. Muse & K. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52-77). New York: The Guilford Press.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Neff, J., Ballesteros, F., Dafouz, E., Martínez, F., & Rica, J-P. (2007). A contrastive functional analysis of errors in Spanish EFL university writers' argumentative text: corpus-based study. In E. Fitzpatrick (Ed.), *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse* (pp. 157-165). New York: Rodopi.

Nemser, W. (1971). Approximative systems of foreign language learners. *International Review of Applied Linguistics in Language Teaching*, 9(2), 115-123.

- Nerlich, B., & Clarke, D. D. (2003). Polysemy and flexibility: introduction and overview. In B. Nerlich, Z. Todd, V. Herman & D. Clarke (Eds.), *Polysemy: Flexible patterns of meaning in mind and language* (pp. 3-30). Berlin: Mouton de Gruyter.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. H. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 125-152). Amsterdam: John Benjamins.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing Company.
- No, G., & Chung, T. (2006). Multiple effects and the learnability of English unaccusatives. *English Teaching*, 61(1), 19-39.
- Obler, L. K., & Goral, M. (2007). Bilingual lexica. In G. Jarema & G. Libben (Eds.), *The mental lexicon: core perspectives* (pp. 185-205). Oxford: Elsevier.
- O'Dell, F. (2015). Creating new words: affixation in neologisms. *ELT Journal*, 70(1), 94-99.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- Oshita, H. (1997). "The unaccusative trap": L2 acquisition of English intransitive verbs. Unpublished doctoral dissertation. University of Southern California.
- Oshita, H. (2000). What is happened may not be what appears to be happening: a corpus study of 'passive' unaccusatives in L2 English. *Second Language Research*, 16(4), 293-324.
- Pae, H. K., Schanding, B., Kwon, Y-J, & Lee, Y-W (2014). Animacy effect of language specificity: Judgement of unaccusative verbs by Korean learners of English as a foreign language. *J Psycholinguist Res*, 43, 187-207.
- Panah, M. A., & Padakannaya, P. (2008). Phonological and orthographic processing skills in children learning to read and spell Persian and English concurrently. *Asia Pacific Journal of Speech, Language and Hearing*, 11(3), 161-173.
- Park, Hyeson. (2015). Actually and in fact in a learner corpus of English. *The Journal of Linguistic Science*, 73, 81-102.
- Park, Hyeson. (2016). Adverbs in a learner corpus: focusing on predicational adverbs. *Studies in Modern Grammar*, 87, 185-213.
- Park, K-S (2005). *L2 acquisition of English unaccusative verbs by native speakers of Korean*. MA thesis. Southern Illinois University Carbondale.
- Park, Keunhyung, & Choi, Incheol. (2016). The acquisition of English negatives by young Korean speakers: a learner corpus study. *Studies in Modern Grammar*, 89, 169-187.

- Park, Myungeun. (2015). *An analysis of spelling errors in high school students' dictation*. MA thesis. Inha University.
- Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading*, 7, 3-24.
- Perlmutter, D. (1978). Impersonal passives and the unaccusative hypothesis. *Proceedings of the Berkeley Linguistics Society* 38. 157-189.
- Pollock, J. J., & Zamora, A. (1983). Collection and characterization of spelling errors in scientific and scholarly text. *A Journal of the American Society for Information Science*, 34(1), 51-58.
- Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23-38.
- Poulisse, N. (1999). *Slips of the tongue: speech errors in first and second language production*. Amsterdam: John Benjamins Publishing Company.
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1-39.
- Prat-Sala, M., Shillcock, R., & Sorace, A. (2000). Animacy effects on the production of object-dislocated descriptions by Catalan-speaking children. *Journal of Child Language*, 27, 97-117.
- Purvis, C. J., McNeill, B. C., & Everatt, J. (2016). Enhancing the metalinguistic abilities of pre-service teachers via coursework targeting language structure knowledge. *Annals of Dyslexia*, 66, 55-70.
- Rajab, A. S., Darus, S., & Aladdin, A. (2016). An investigation of semantic interlingual errors in the writing of Libyan English as Foreign Language learners. *Arab World English Journal*, 7(4), 277-296.
- Ramirez, G., Chen, X., Geva, E., & Kiefer, H. (2010). Morphological awareness in Spanish-speaking English language learners: within and cross-language effects on word reading. *Reading and Writing*, 23, 337-358.
- Ravn, K. E., & Gelman, S. A. (1984). Rule usage in children's understanding of "Big" and "Little". *Child Development*, 55(6), 2141-2150.
- Reed, D. K. (2012). *Why teaching spelling?* Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Rhee, S. C., & Jung, C. K. (2012). Yonsei English Learner Corpus (YELC). *Proceedings of the First Yonsei English Corpus Symposium*. Seoul, Korea. 31 March 2012, pp 26-36.

- Rhee, S. C., & Jung, C. K. (2014). Compilation of the Yonsei English Learner Corpus (YELC) 2011 and its use for understanding current usage of English by Korean pre-university students. *Journal of the Korea Contents Association*, 14(11), 1019-1029.
- Richards, J. C. (1974). A non-contrastive approach to error analysis. In J. C. Richards (Ed.), *Error analysis perspectives on second language acquisition* (pp. 172-188). London: Longman.
- Ringbom, H. (1983). Borrowing and lexical transfer. *Applied Linguistics*, 4(3), 207-212.
- Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters Ltd.
- Ringbom, H. (2001). Lexical transfer in L3 production. In J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives* (pp. 59-68). Clevedon: Multilingual Matters.
- Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters Ltd.
- Roach, P. (2000). *English phonetics and phonology: a practical course*. Cambridge: Cambridge University Press.
- Römer, U. (2004). Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In G. Auston, S. Bernardini & D. Stewart (Eds.), *Corpora and language learners* (pp. 151-168). Amsterdam: John Benjamins.
- Römer, U. (2006). Pedagogical applications of corpora: some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 121-134.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Ryu, Miryeong, & Park, Mae-Ran. (2017). A comparative analysis of coordinator ‘and’ between Korean learner corpus and ICE. *Journal of Language Sciences*, 24(2), 79-99.
- Salam, S. (2016). English spelling errors of Arab students at tertiary level: A serious concern. *International Journal of Arts and Sciences*, 7(5), 205-210.
- Salem, I. (2007). The lexico-grammatical continuum viewed through student error. *ELT Journal*, 61(3), 211-219.
- Sapir, E. (1925). Sound patterns in language. *Language*, 1(2), 37-51.
- Schachter, J., & Celce-Murcia, M. (1977). Some reservations concerning error analysis. *TESOL Quarterly*, 11(4), 441-451.

- Schlagal, B. (2002). Classroom spelling instruction: History, research and practice. *Reading Research and Instruction*, 42(1), 44-57.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word association and verbal suffixes. *SSLA*, 20, 17-36.
- Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145-171.
- Schreuder, R. & Weltens, B. (1993). The bilingual lexicon: An overview. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 1-10). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Scott, M. (2012). *WordSmith Tools*. Liverpool: Lexical Analysis Software.
- Seidlhofer, B. (2005). English as a lingua franca. *ELF Journal*, 59(4), 339-341.
- Selinker, L. (1972). Interlanguage. *IRAL*, 10(3), 209-231.
- Shalaby, N., Yahya, N., & El-Komi, M. (2009). Analysis of lexical errors in Saudi college students' compositions. *Journal of the Saudi Association of Languages and Translation*, 2(3), 65-92.
- Shimotori, M. (2013a). *Conceptual contrast of dimensional adjectives in Japanese and Swedish: Exploring the mental lexicon by word-association test*. PhD dissertation. Umea University.
- Shimotori, M. (2013b). Conceptual contrast of dimensional adjectives in Japanese and Swedish: Exploring the mental lexicon by word-association test. In A. Saxena & A. Viberg (Eds.), *Multilingualism: Proceedings of the 23th Scandinavian Conference of Linguistics* (pp. 296-304). Uppsala: Acta Universitatis Upsaliensis.
- Shin, Jeong-Ah. (2011). Overpassivization errors in Korean college students' English writings. *Korean Journal of Applied Linguistics*, 27(3), 255-273.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London and New York: Routledge.
- Song, Sanghoun. (2018). Phraseological analysis of learner corpus based on language model. *Language and Information*, 22(1), 123-152.
- Steen, G. (2007). Finding metaphor in discourse: Pragglejaz and beyond. *Culture, Language and Representation*, 5, 9-25.
- Stefanowitsch, A. (2006). Words and their metaphor: A corpus-based approach. In A. Stefanowitsch & S. Gries (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 63-105). Berlin: Mouton de Gruyter.

- Stein, G. (2007). *A dictionary of English affixes: their function and meaning*. Munich: Lincom Europa.
- Tarone, E. (1981). Some thoughts on the notion of communication strategy. *TESOL Quarterly* 15(3), 285-295.
- Tarone, E., Cohen, A. D., & Dumas, G. (1983). A closer look at some interlanguage terminology: a framework for communication strategy. In C. Færch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 4-14). London: Longman.
- Taylor, J. R. (1995). *Linguistic categorization*. Oxford: Oxford University Press.
- Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: insights from an error-tagged EFL learner corpus. *The Modern Language Journal*, 97(1), 77-101.
- Thomas, M. S. C., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language*, 43, 44-66.
- Thomas, M. S. C., & Van Heuven, W. J. B. (2005). Computational models of bilingual comprehension. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 202-225). Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Tono, Y. (2002). *The role of learner corpora in SLA research and foreign language teaching: The multiple comparison approach*. PhD dissertation. Lancaster University.
- Tono, Y. (2004). Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In G. Auston, S. Bernardini & D. Stewart (Eds.), *Corpora and language learners* (pp. 45-66). Amsterdam: John Benjamins.
- Tucker, G. H. (1999). *The lexicogrammar of adjectives: A systemic approach to lexis*. London: Cassell.
- Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language*, 28, 649-667.
- Varadi, T. (1983). Strategies of target language learner communication: message adjustment. In C. Faerch & G. Kasper (Eds.). *Strategies in interlanguage communication* (pp. 79-99). London: Longman.
- Walker, C. (2008). *A corpus-based study of the linguistic features and processes which influence the way collocations are formed*. Unpublished PhD thesis. University of Birmingham.

- Wang, M., Cheng, C., & Chen, S-W. (2006). Contribution of morphological awareness to Chinese-English biliteracy acquisition. *Journal of Educational Psychology*, 98(3), 542-553.
- Wang, M., & Geva, E. (2003). Spelling performance of Chinese children using English as a second language: Lexical and visual-orthographic processes. *Applied Psycholinguistics*, 24, 1-25.
- Wang, Ying. (2016). *The idiom principle and L1 influence: A contrastive learner-corpus study of delexical verb + noun collocations*. Amsterdam: John Benjamins Publishing Company.
- Weinreich, U. (1968). *Languages in contact: findings and problems*. The Hague: Mouton Publishers. (Original work published 1953).
- Westwood, P. (2014). *Teaching spelling: Exploring commonsense strategies and best practices*. Oxon: Routledge.
- White, L., Brown, C., Bruhn-Gravito, J., Chen, D., Hirakawa, M., & Montrul, S. (1999). Psych verbs in second language acquisition. In E. C. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach* (pp. 171-196). Amsterdam: John Benjamins Publishing Company.
- Yilmaz, F. (2014). Turkish as a foreign language learners' awareness and use of morphology in guessing the meanings of unknown words from context: A case study. *Educational research and Reviews*, 10(5), 675-683.
- Yip, V. (1995). *Interlanguage and learnability: From Chinese to English*. Amsterdam: Benjamins.
- Zheng, L., & Xiao, R. Z. (2015). A corpus-based study of collocation in Chinese EFL learners' oral production. *Corpus Linguistics Research*, 1, 83-108.
- Zimmermann, R. (1986). Classification and distribution of lexical errors in the written work of German learners of English. *Papers and Studies in Contrastive Linguistics*, 21, 31-40.
- Zimmermann, R. (1987). Form-oriented and content-oriented lexical errors in L2 learners. *IRAL*, 25(1). 55-67.
- Zobl, H. (1989). Canonical typological structures and ergativity in English L2 acquisition. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 203-221). Cambridge: Cambridge University Press.
- Zughoul, M. (1991). Lexical choice: toward writing problematic word lists. *IRAL*, 29(1). 45-60.

Online Dictionaries

Collins Online English Dictionary. Retrieved from
<http://www.collinsdictionary.com/dictionary/english>.

Merriam-Webster Online English Dictionary. Retrieved from <http://www.merriam-webster.com/dictionary>.

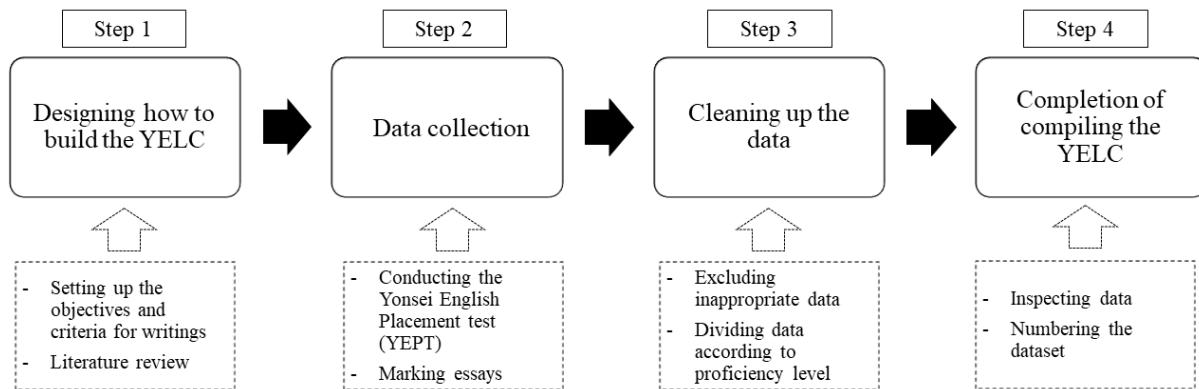
Standard Korean Language Dictionary. Retrieved from <http://stdweb2.korean.go.kr/main.jsp>.

Appendices

Appendix 1: The general information of the YELC (Rhee & Jung, 2014)

Note: The YELC was compiled through the Yonsei English Placement Test (YEPT), which measures the English speaking and writing proficiency level of new students. The YEPT is a computer-based test, where the test-takers should perform both speaking and writing tasks in the computer lab. The writing sections of the YEPT became the source of the YELC, where the test-takers should write narrative or argumentative essays.

1) The compiling process of the YELC



2) The statistics of the YELC

	Narrative section	Argumentative section	Overall
Number of texts	3,286	3,286	6,572
Number of token	315,317	770,511	1,085,828
Number of type	11,308	16,416	21,839
Standardised TTR⁴⁰	73.38	76.79	75.93
Number of sentences	25,386	52,814	78,200
Average number of words in a sentence	12.36	14.57	13.85

⁴⁰ According to Rhee & Jung (2014), they adjusted the standardised TTR basis in WordSmith Tools to its lowest setting at 50.

Appendix 2: The data samples of sub-corpora in the YELC

Note: The YELC consists of two genres: narrative and argumentative essays. The test-takers are advised to write a narrative essay of around 100 words and an argumentative essay of around 300 words within 60 minutes. The YELC is a raw corpus, which is not annotated with any type of tagging. Here are some samples of essays from each level.

1) Sample essays from the low level

Text: 289_01
I like biology and chemistry in high school. the reason why I like those subject is because those subject is so interest to me for example, few moles can react to bigger one and in real there are just water but in chemistry water is H ₂ O which is consist of oxygen and hydrogen. doesn't it exciting? in biology, I liked a part of genetic the part is so interest to me because I never been study about this area in class, the time when I was so exciting is experiment time in experiment, there are fantastic things for instant teacher dye cells to see it and my blood react to anti O blood type because of these reason I loved this two subject in my high school

Text: 289_02
I think that all korean men be forced to complete military service because the time when we are living is war time I think we should go military even by force but military must have many special part all human have a few expert part than the others so military should separate correct people in correct part not just separate people just by number before going to military men should take more test not just body test for example solve test will find genious about solve password which the enemy use or shooting test to find best shooter or I write about what happen when korean are not forced to complete military service frankly I think we are living now because USA military are in my country and former korean men are forced to complete military service i heard about story that north korea kid are trained to be soldier since middle school but our country's middle school student are tired in study, never get a train to be soldier before go to military this situation, which side will be winner when korea war arise again? I can't sure our country will destroy the enemy and reunify north and south

2) Sample essays from the intermediate level

Text: 2795_01

The subject I like most is the korean. there are two reasons First, through korean I can learn many poems and novels. It increase qualiy of my life and I can enjoy the literatures through this education. It is very important not only for my study ability but also for my life. Second, I can write essay well through writting education in korean class. Writting is very hard for me but through this class I can do better and I can enjoy the writting. I think It is meaningful for me as It increase the quality of my life.

Text: 2795_02

There are many controvertial points about the physical punishments. somebody think it can be very harmful for growing children. but I agree with the physical punishment. I think if the punishment is used right, It can be not only effective way for controlling children but good means for educating children. so There are three big reasons. Frist, the physical punishment is most effective way for controling the children. It is very hard for teachers to make children follow the rules, if there are no physical punishments. So, teachers should choose other punishment which can be worse for children than physical punishments. So, I think the physical punishments are most usual and effective way for controlling childrean. Second, phisical punishments can teach children that following the rules is important and if they break the rules, they should get the punishments. In the society, civilians are watched by law and if they break the law, they get punishments by police. But children are free from law so, in the school teachers should show the children that rule is important through some powerful means. and I think phisical punishments are the most effective way. Third, it can be good education for childrean if it is based on the proper rules. For example, my middle school english teacher always did physical punishments but nobody complained it. Because she always explained the rules and we agreed that rules. Although we were scared of the physical punishment, we could learn the rule and how important the rules are. so, I think if children agree the punishments, physical punishments can be a good education. Because of these three reasons, I think the physical punishmets are good way for teaching children.

3) Sample essays from the advanced level

Text: 1329_01

I usually spend saturday playing with my pet. I raise two pandamice for my pets. They got their name 'panda' because they have black spots like panda around their neck. I usually play with them training some tricks. They understand what I order and follow it. It's so happy to be with them. After i play with them, I watch American dramas. Thesedays I'm watching 'Smallville', which is about superman's highschool days. So interesting and thrilling it is, i once watched 6 series of episodes of Smallville one night. Lastly, I check my e-mail and twitter before I go to sleep.

Text: 1329_02

Thesedays as more and more people start smoking, also the number of people who smoke in buildings rised. This is because people want to smoke in warm and convenient places or don't want to bother going out of building taking an elevator or stairs. However smoking in buildings may cause plenty of serious problems. These are the reasons why smoking should be banned in all public buildings. First of all if one is smoking, the smoke from the cigarette diffuses to all area and affects other people. If this place is in a building, the problem gets worse. The smoke is locked in certain zone, so it can't be spread to a larger area. The locked smoke circulates the closed area and is inhaled by people who didn't even touch a cigarette. Inhaled cigarette smokes contain more of harmful elements like nicotine and tar, because they are not filtered by the cigarette. Those people who are affected by cigarette smokes are called 'second hand smokers'. The smokes destroy second hand smokers' lungs without their will, and finally brings lung cancer. Second, smokers in public buildings throw away their cigarettes anywhere on the floor. It causes sanitary problems like spreading of flu virus. Moreover the cigarette's unextinguished flame leaves black and gray marks on the ground. It's hard to remove them because they are made by hot flames. Those marks make people who use the building feel that the building is not quite a clean place and give visitors negative images. Lastly, although some smokers put their used cigarettes in a trash can, they leave ashes on the ground during smoking. That behavoirs raise the possiblility of fire, which brings about death of lots of innocent people in the building. For those reasons, people shouldn't even think of smoking inside and smoking in buildings must be banned by law, for everyone's sake.

Appendix 3: The additional information for collocational errors of dimensional adjectives

1) Error occurrence rate for collocational errors according to proficiency level

Dimensional adjectives	Low Error / Correct Error rate (%)	Intermediate Error / Correct Error rate (%)	Advanced Error / Correct Error rate (%)
<i>large</i>	5 / 21 (23.8%)	13 / 65 (20.0%)	1 / 6 (16.7%)
<i>small</i>	17 / 76 (22.4%)	41 / 249 (16.5%)	1 / 15 (6.7%)
<i>high</i>	5 / 82 (6.1%)	12 / 309 (3.9%)	1 / 19 (5.3%)
<i>low</i>	7 / 45 (15.6%)	16 / 112 (14.3%)	1 / 9 (11.1%)
<i>long</i>	0 / 87 (0.0%)	2 / 332 (0.6%)	1 / 20 (5.0%)
<i>short</i>	0 / 47 (0.0%)	3 / 137 (2.2%)	1 / 9 (11.1%)
Total	34 / 358 (9.5%)	87 / 1,204 (7.2%)	6 / 78 (7.7%)

2) Example concordance lines of collocational errors of dimensional adjectives

Interlingual errors
<p>Many people's <i>convenience</i> is large than your life. I am sorry about But some teachers give children large punishment such as to hit second, in world there are large winds like trade wind. that But <i>disadvantage</i> is so large that it can overwhelm the</p> <p>Moreover, if we have to take sentence may hurt our mind. A small or big <i>fight</i> with other country And that <i>danger</i> is much small small smaller than that of people who are</p> <p>study's <i>effectiveness</i> was low. But when I studied math, I making <i>possibility</i> of accident low, everything that makes driving stronger than women. After a long discipline, women might get ([sic], sentences from the YELC)</p>

Intralingual errors

animal in small experiments to study every thing again with and china. their contries have

In this situation, even a are no adult, they can lose number of automobile accidents.

Another students have much students with their affection, our netizenship will become

to teachers. Moreover, the If one teacher's autority is today young's health is very

cast7e(?). I heard it is very

These are just my

large
large
large

small
small
small

high
high
higher

low
low,
low

long

short

experiment. Altough I don't know effort and time after completing military so for defense our

murmur could disturb your self-esteem than adults so i habits like not using cellular

violation. The punishment often personality and charisma, not than before. And through these

dignity of teachers is related students don't listen their for irregular eating and

and excellent. When I come back

opinion, but I think there

([sic], sentences from the YELC)

Appendix 4: The additional information for over-passivisation errors of non-alternating unaccusative verbs

1) Example concordance lines of over-passivisation errors

Type 1: typical passive errors
<p>person's privacy could not be existed anymore and his or her personal that this kind of village is existed in Seoul. I want to visit here many assault and harassment is existed in Internet, so some people facts, medical experiments are existed for curing human, not for</p> <p>driving, car accident would be happened very much. When drivers of these behaves are happened because there is no way to if the chance of camp is happened to me, i would like to join the most sad thing in my life was happened My middle school friend died.</p> <p>so, traffic accidents that is occurred by phoning driver will decrease. that more accident can be occurred by using cell phone while you have hardship that is occurred in social life. she doesn't just and many bad events are occurred in schools. For example, we can</p> <p>Although North korea are remained many men go to army is not good that this kind of punishment is remained is this is the most certain way I think that internet must be remained as anonymous. Because anomimity rights of teachers cannot be remained and this situation will tend to</p> <p>maybe insist that privacy is appeared. But if you worry offending Internet, many people can be appeared by ways such IP searching. And our opinions and It could be appeared by using fake name. Being including this problem has been appeared so often. But as these problem</p> <p style="text-align: right;">([sic], sentences from the YELC)</p>

Type 2: BE + base form errors
<p>Korean and Japan must be exist and I want to experience them. physical punishment has to be exist, and it is irreplaceable, because children in schools should be exist to help the students. If my best friend were not exist, I would not enter the Yonsei</p> <p>If that situation is happen, the punishment has competent An explain, that would not be happen. Of course when I knew what I do that situation wouldn't be happen. And through physical punishment, And those are all happen. Because they use nickname on the</p> <p>claim that shouldn't be occur and that's sort of abusing traffic accidents are usually occur while handling with two hands. "It will be not occur to me, it was some story to This is can be occur whenever, wherever. For example,</p> <p>addition to, many reasons are remain. but i think that these three way to use. But, It must be remain for students. military service should be remain for political needs. The address or other evidents are remain in the server. Maybe I would be</p> <p>And there, big problem isn't appear which is considered. Rather,</p> <p style="text-align: right;">([sic], sentences from the YELC)</p>

Type 3: transitive errors carrying an object

<p>Though there can also a dangerous of naturally if not sections, so it's necessary to a horror in mind. And not with this peninsula and can fight with my friends or when I or her student. But it can when the North Korea walked down in stairs. when we Jeju island alone. before i It made me nervous. when i and meet my friend. when I phones while driving can of family. But these can children's health. Smoking can This situation often can people and mental punishment illegal action. The punishment declined, they didn't want to between two desires willing to using our real name let us our creating name, so it can disappointed and they do not worse and worse and will</p>	<p>exist some negative points about the my one hand. But never using a the military institution. But if a way to mend a bad action from happen war in this area. happen bad works everybody doesn't look happen some problems like student's happen the problem. If urgent situation arrived first floor, one of my friends arrived jeju island, i was very nervous arrived Singapore, its temperature is arrived New Zealand I can see my friend occur serious accidents. When drivers occur another accidents. In roads, occur many kind of breath disease. occur automobiles accident and people remains more problems in children than remains stress. So, the students release remain a teacher. So they will abandon remain their power for reign people appear our opinion less frequently. If appear ourselves more accurately and appear media. However, if we use real appear big problem in our society. In</p>
---	---

([sic], sentences from the YELC)

2) Error occurrence rate for individual unaccusative verbs according to proficiency level

Unaccusative verbs	Low Error / Correct (Error rate (%))	Intermediate Error / Correct (Error rate (%))	Advanced Error / Correct (Error rate (%))
<i>exist</i>	16 / 58 (27.8%)	39 / 147 (26.5%)	0 / 9 (0.0%)
<i>happen</i>	24 / 133 (18.0%)	53 / 452 (11.7%)	1 / 29 (3.4%)
<i>arrive</i>	2 / 3 (66.7%)	6 / 40 (15.0%)	0 / 1 (0.0%)
<i>occur</i>	19 / 46 (41.3%)	73 / 186 (39.2%)	2 / 10 (20.0%)
<i>remain</i>	16 / 29 (55.2%)	37 / 98 (37.8%)	1 / 8 (12.5%)
<i>appear</i>	6 / 16 (37.5%)	8 / 57 (14.0%)	0 / 3 (0.0%)

Appendix 5: The additional information for derivational morphological errors

1) Examples of derivational morphological errors

	Derivational affix	Examples of errors
Class-changing (Suffix)	<i>-able</i>	importable, unfocusable, problemable, hurtable, unchoicable, handable, terrible, benefitable, memorable, irreversible, permitable, inevitable, nonprofitable, unexpected, needable, communicatable, available, suggestable, inendurable, unperceptionable, passionate, effectable, satisfiable, speechable
	<i>-ible</i>	relaxible, incomparable, inversatible, inevitible, acceptable
	<i>-er</i>	glober, crimer, interver, aboarder, drunker, educator, advicer, conducter, consealer, creater, illuster, instructer, legister, policer, protecter, sculpter
	<i>-less</i>	meanless, thinkless, taughtless, rudeless, washless, licenseless, lackeless, quitiless, carefuless
	<i>-ish</i>	Iranish
	<i>-ly</i>	objectly, misdiscernly, strately, unkownly, outly, insidely, espectually, continely, activitly, oftenly, consistly, resultingly, slicely, indeedly, cursely, violately, smally, purposedly, realitly, speedly, diversitely, conclusionly, consequencely, detail, encourageously, enoughly, expandly, forcely, hopely, instinctly, mistakenly, restrictly, ruinly, selfly, thinkfuly, usely, worsely, zegzagly
	<i>-ness</i>	faultness, obeyness, sleepness, triness, sharpness, harmness, careness, advantureness, afraidness, angriness, carefulessness, cutiness, fearness, frightness, mindness, onesideness, saveness, shameness, sorrowness, thinkness, urgentness
	<i>-ize(is)e</i>	paralize, abolise, automize, diminise, differenize, annoise

Class-maintaining (Prefix)	<i>un-</i>	unadequate, unallowing, unchoiable, unconvinient, undirect, uneffective, unequality, unfavor, ungood, unhumane, unidentity, unjustice, unproper, untemporary, unrational, unwantedly, unvividly, unviolentic, unsmoking, unseriously, unrightly, unpropound, unpleasure, unmeaningful, uninnoent, unethic, undirective, unconsiderate, unconcord, uncomport, unbanned, unagree
	<i>in-, im-, il-, ir-</i>	invision, inexpectable, inconcentration, incognitive, inethical, ineducative, ingeneously, inpolite, inrude, irregal
	<i>dis-</i>	disconcentrate, disencouraging, diskind, discorrect, disgrow, disworth
	<i>re-</i>	rehappen, reexperiment, recorreted, reprobability, restrict, recomeent, remisbehaving

2) Examples of derivational morphological errors according to error type

Type	Examples of errors
A	terrible, relaxible, educater, irreversable, suggestable, counselor, sculter, protecter, possesser, instructer, creater, conductor, acceptable, inevitable
B	crimer, obeyness, shapeness, unadequate, uncorrect, unactively, undirect, unjustice, unidentity, unconvinient, uneffective, unequality, unhumane, unallowing, unfocus, unresponsibility, unrational, unproper, inconcentration, incognitive, diskind, disconcentrate, discorrect, uneviable, undispensable, inendurable, carefulness, urgentness, embarassness, awakeness
C	meanless, purposedly, violately, resultly, conclusionly, diversitely, speedly, continuely, cursely, consistly, resultingly, sleepness, triness, faultness, unmeaniful, disencouraging, memoriable, permitable, legister, illuster, advicer, forcely, hopely, mistakenly, easely, repeatly, paradoxly
D	unchoicable, unfocusable, problemable, importable, hurtable, handable, drunker, thinkless, taughtless, rudeless, insidely, indeedly, oftenly, slicely, outly, objectly, misdisernly, smally, harmness, ungood, unvariably, untemporary, unfavor, unsing, unsubmitted, invision, inexpectable, disgrow, disrespection, disworth, communicatable, effortable, passionate, inspeeckable, benefitable, inversatible, washless, rudeless, unwantedly, unpurposely, worsely, worldwidey, unnoticedly, detaily, mindness

3) Examples of interlingual influence in Type D errors

Level	Examples of errors
Low	hurtable, problemable, unchoicable, unfocusable, taughtless, insidely, smally, indeedly, slicely, outly, disworth, disrespecton, inexpectable, unchoicable, unfocusable
Intermediate	communicatable, co-existable, benefitable, thinkless, licenseless, zegzagly, detaily, ingeneously, irrestorable, rehappen, reexperiment, recorrected, uncarefulness, unpurposely, unrightly, uncomfortability
Advanced	wordly

Appendix 6: The additional information for orthographic errors

1) Examples of interlingual spelling errors

Type	Examples of errors
[p]-[f] substitution	fafa, fationate, fension, feriod, flantinum, fragnancy, helf, helfed, imfortant, liverfool, resfect, sofrano appect, cellpones, comportable, compotable, discomfortable, mapia, parpomence, perfect, perfectly, plag, pones, punction, pysical, pysically, replex, specipic, uncomportable
[l]-[r] substitution	congraturation, curture, itary, groomy, probrem, sordiers, abroader, aralm, archols, arive, congraturate, congraturated, curiculrum, deveromponts, irregal, lovery, loyar, neckress, razy, rebra, resson, ribrary, rocated, uproad, uproading ploblem, conglatulated, halmful, lainy, ploblems, reglet, sideload, singapole, sulf, sulflng
[b]-[v] substitution	liverty, obiously, faborite, lavoratory, mablous zentleman

2) Examples of intralingual spelling errors

Type	Examples of errors
Inflectional	abondonded, activitys, allowd, answerd, anythings, applyed, applyer, baned, beated, bomming, carring, celebirty, cellphons, childeeren, choosen, complained, concernd, considerd, controled, controlls, counseled, destroyiong, developed, devicies, disapeارد, distroyed, disturbes, drinked, earlyer, easyily, eated, eaten, eatting, envys, exitting, fighted, forbiding, forgived, franklly, frozen, frightend, groundied, hanppend, happend, hateing, healther, heared, helpping, hided, hitted, humen, implys, japaness, laughes, leardned, lended, limitted, listenning, luckilly, maded, mankinds, memberes, militarys, monkies, nameing, nobodys, nowaday, obviously, occuerd, olddest, openning, permitted, philosophier, photoes, photographes, physically, quited, regreted, rised, ruunned, secretarys, shoping, sking, speaked, stoped, stronggest, studing, studied, submited, taked, teached, teachs, teared, thoughted, trys, uncontroled, undergos, understanded, useed, useing, visisted, wantted, weared, worrys, writting

Derivational	absolutly, activitly, additionaly, anonymousity, appliable, appropriatly, beautyfull, carful, chiness, christion, contemporay, conveinece, courageous, cultual, dangerouse, easilly, educaion, efficiest, enginer, entirly, especaily, excitemant, especiaiy, experimnets, faithfullness, famouse, forcful, freedon, globalizaiton, historicaly, historican, honestely, humorus, illegally, informaion, invalueable, irritaion, japanes, joyfull, managable, nationall, officialy, organizaion, personallity, personnal, possiblty, professionaly, punishemt, quantitiy, question, reallity, reasonble, regulary, responsiblility, responsilbe, rudly, sadely, safty, secretely, seriously, shortlly, similary, succesion, traditionl, trafffic, unbelieveble, unchangable, uncureable, variaty, varioous
--------------	---