Lexical Issues in L2 Writing

Lexical Issues in L2 Writing

Edited by

Päivi Pietilä, Katalin Doró
and Renata Pípalová

Lexical Issues in L2 Writing

Edited by Päivi Pietilä, Katalin Doró and Renata Pípalová

This book first published 2015

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

TABLE OF CONTENTS

Table of Contents

Part III Collocations and Lexical Bundles

INTRODUCTION

The inspiration for this book was the ESSE conference (European Society for the Study of English) held in Košice, Slovakia, in August 2014, where the three of us convened a session bearing the same title, *Lexical Issues in L2 Writing*. We decided to write up our presentations in the form of book chapters, and to invite other scholars interested in similar issues to join us in the project. The result is this volume, dedicated to research into various lexical aspects of second language writing. The authors of the chapters are experienced scholars who share a genuine interest in matters lexical, particularly as manifested in the written performance of language learners. Our aim was to produce a state-of-the-art presentation of current views and recent research on vocabulary acquisition and use in a second or foreign language.

Lexis enjoys a special status in any language, in that it undergoes change more rapidly than grammar, which tends to be fairly stable. Indeed, lexis has to sensitively reflect real-life developments and keep abreast with the diverse communicative needs of the respective communities of practice. At the same time that new words keep emerging, those no longer used gradually become obsolete and disappear from the lexical system. It is also frequently attested that word meanings are susceptible to change, whereby their senses may be narrowed, widened or diversely transformed in specialized contexts. Not surprisingly, then, even the mental lexicon of native speakers is subject to continuous development over its lifespan, being gradually enriched, or otherwise.

Achieving native-like command of second language vocabulary poses a real challenge. It may well be easier to master a system of rules, such as the grammar of a language, than an ever-growing class of lexical items. What is more, it is not merely the size of the L2 mental lexicon that matters, but also the appropriate use of the words that one has access to. In fact, knowing a word entails a number of sub-skills, from being familiar with its spoken or written form to knowing its synonyms, grammatical functions, and other characteristics, as well as knowing how to use it appropriately (e.g., Nation, 2001). In particular the acquisition of collocations has been shown to be difficult even for the most advanced learners (Fan, 2008; Laufer & Waldman, 2011).

The vocabulary of a language is sensitive to a wide range of co-textual and contextual considerations. Thus, in idiomatic use, not only does a lexical choice have to be grammatically fitting; simultaneously, it has to be appropriate in style and register, and in a number of other respects. Indeed, words enter into a myriad of relationships: for instance, they can combine appropriately only with particular items in collocations or bundles, they enter into numerous cohesive chains, have particular currencies, may become fashionable or obsolete. Moreover, words may differ almost imperceptibly in shades of meaning, bear various connotations, and invoke distinct cultures. They may be charged with evaluative potential and radically change the tone or formality of a passage.

With all this in mind, we strove to grasp various aspects of this multifaceted topic. In order to show its comprehensiveness, we decided to pursue the principle of unity in diversity. We invited authors coming from a number of schools of thought and from different language backgrounds. The common interest bringing the authors together is naturally lexical; more specifically, L2 lexis in authentic use. The studies, however, have grown out of a much wider array of disciplinary backgrounds. Although most chapters are rooted in second language acquisition, a number of other branches of linguistics are either drawn on directly or at least implicated secondarily. The list includes corpus linguistics, English for academic purposes, (academic) writing pedagogy, stylistics, text linguistics, discourse analysis, pragmatics, psycholinguistics and sociolinguistics. In fact, disciplinary diversity has been among our priorities.

The present volume deals exclusively with lexis in writing. Naturally, interaction with readers differs conspicuously from oral communication, whether with a single interlocutor or an audience; it is marked by certain characteristic features. Academic writing comes into existence as a result of the writing process; this in turn involves numerous stages, ranging from planning and the earliest drafts to the final, edited, fixed, publishable product, although the interim stages and breaks in the process are not made visible. Despite the seemingly monologic nature of written discourse, the author has to take into account and collaborate with the prospective audience, facilitating and enhancing their perception of coherence. This may be achieved in a number of ways: by employing appropriate structure/organization with paragraphs and sections, by weaving in a web of cohesive links and chains, and by creating a smooth information flow. It is perhaps needless to add that lexical choices also rank among the prominent features which may either enhance or impair the perception of coherence, since they are indispensable for the negotiation of meaning.

All the chapters in this volume are corpus-based, exploring authentic data gathered in very recent corpora. Some authors have made use of ready-made corpora, such as SUSEC (Stockholm University Student English Corpus); others have employed corpora tailor-made to suit their particular research designs. Understandably, such corpora were not decontextualized, but rather compiled with a knowledge of their "communicative function in the community in which they arise" (Sinclair, 2005). Most authors analysed writing by their local non-native populations, although one study compares non-native writing across two L1 language backgrounds (Pietilä). Lexical patternings were examined in a multitude of L2 discourses. The genres scrutinized include free compositions, essays, portfolios, BA theses, MA theses, and monographs. The chapters also cover various text types and fields, including for example both expository and argumentative prose, and dealing, among a variety of topics, with writing on both literature and linguistics. The writing under investigation has all been produced in L2 English, except for one study which focuses on writing in L3 French (Mutta).

A clear majority of the studies investigate lexis in writing associated in one way or another with educational contexts, spanning the upper-secondary (Henriksen & Danelund) and undergraduate levels (the majority of chapters). This indicates that the authors had pedagogical implications in mind, and some even addressed them explicitly. Two chapters explore the writing of subjects beyond the scope of formal education; one looks at writing by in-service teacher trainees (Lehmann), the other examines professional academic writing by scholars (Pípalová). In other words, the writers in the studies vary in several respects, including L1 background, age, degree of proficiency, education, and erudition.

Most of the authors found it useful to carry out their research using modern, computer-assisted tools, such as VocabProfiler (Cobb, n.d.), RANGE (Nation, n.d.) or AntConc (Anthony, n.d.). These instruments enable the researcher to process huge quantities of data, and facilitate the comparability of findings across various studies. Occasionally, however, manual data collection proved necessary, chiefly due to a qualitative focus in the research. Whatever the approach adopted, quantitative results were often matched by qualitative analysis, carefully interpreting and contextualizing the findings.

Some of the authors compared related written discourse produced by native and non-native writers (Erman; Pípalová); others measured the performance of L2 writers by L1 norms implicitly, using the yardstick of various service lists established by processing huge native corpora, such as the NAWL, AWL, or NGSL (Doró; Henriksen & Danelund; Lindgrén;

Lehmann); still others combined the two approaches (Mutta; Pietilä). Some authors also compared their results for non-native writers with studies dealing with native discourse (Lehmann).

The volume addresses a multitude of lexical aspects, including – to name but a few – lexical frequency, lexical density, lexical distribution, lexical richness, lexical variation, lexical diversity, lexical sophistication, and lexical errors. It should be noted, however, that these terms may not be always used in the same sense, which follows from the diversity of the epistemological traditions represented in the volume. In addition to the variety of lexical aspects, numerous vocabulary strata were subjected to analysis, including academic vocabulary, hedges, boosters, reporting verbs, collocations, and lexical bundles.

The book is divided into three main sections, each approaching lexical issues in L2 writing from slightly different perspectives. The volume opens, however, with a review chapter by **Doró and Pietilä,** in which the writers give an overview of recent developments in research methodology. They look back at the most recent history of research into L2 vocabulary acquisition and use, which has seen the rise of new methods and computer-assisted text analysis tools. They also discuss various computerized vocabulary analysers, prominent service lists established on huge amounts of data, and learner corpora assembled locally or internationally. In addition, the authors ponder the advantages and disadvantages of automated essay scoring and compare it to human rating.

Part I, entitled Influences and Strategies, embraces three chapters, dealing with diverse external influences that affect L2 vocabulary competence.

Henriksen and Danelund raise the crucial topic of the relationship between the size and depth of learners' vocabulary and their writing skills. The chapter reports and synthesizes the results of three previously unpublished studies, each scrutinizing a distinct lexical parameter. All the studies analyse writing by upper-secondary school students. The first study was designed to measure the subjects' receptive vocabulary and their lexical error production; the second aimed at exploring the learners' productive vocabulary size, focusing on lexical variation and sophistication; the last combined receptive and productive vocabulary with a word association task. The chapter reveals a surprising rate of high frequency vocabulary in the students' written production, presumably resulting from avoidance and safe-playing strategies.

Doró discusses differences between two sets of timed argumentative essays by students at the BA level (written several years apart) with respect to three measures: lexical richness, lexical variation, and various

metadiscourse markers (hedges and boosters, together with reporting verbs). She used two software measurement tools, VocabProfiler and AntConc, both of which focus on single-word text parameters. Doró defines lexical richness as the proportion between high and low frequency words, while lexical variation follows from the type/token ratio. The author supplements these single-word parameters with various markers operating at the textual level.

Rather than measuring lexical proficiency, **Mutta** addresses the topic of language transfer, in a study of the influences of L1 and L2 on the L3, in this case French, in a timed writing task arranged at university level, with no recourse to external aids and resources. In her research, supported by VocabProfiler, expository essays were analysed in terms of lexical richness, defined in the chapter as combining the type/token ratio and lexical frequency. Non-native essays are compared to a single native counterpart. The chapter looks at the mutual interaction between various languages activated simultaneously in multilinguals' mental lexicon.

Part II, entitled Disciplinary Differences, consists of three chapters which share several parameters. All of them investigate various lexical features of undergraduate theses; what is more, all are also marked by their cross-disciplinary orientation, the fields in question in all of them being linguistics and literature. The chapters, dealing with different lexical traits, reveal striking lexical discrepancies between the texts of the two fields. While Lindgrén compares theses at the BA and MA levels, the other two chapters focus on MA theses alone. The chapters also differ in their specific research designs and objectives.

Having collected a corpus of MA theses from two non-native and one native group of subjects, **Pietilä** sets out to investigate their conclusion sections as an academic subgenre in terms of numerous lexical parameters. It should be stressed that her non-native subjects differ in their L1 backgrounds (Czech and Finnish), being unrelated in language type and family. The study gives a comprehensive account of lexical aspects. The primary focus of the chapter is on lexical diversity, which is studied by combining intrinsic and extrinsic lexical measures. While the former involve lexical variation and density, the latter explore lexical sophistication, established particularly on the basis of lexical frequency bands. In addition, attention is given to academic vocabulary. While the two non-native groups were found to exhibit comparable lexical traits, differences were more prominent between theses written in the two disciplines, linguistics and literature.

Pípalová explores a corpus of literary and linguistic MA theses written by non-native undergraduates and compares them with published

professional monographs in the two fields, by both native and non-native scholars. The chapter focuses on reporting verbs as part of textual metadiscourse, investigating the corpus with regard to their frequency, distribution and various lexico-semantic features, including verbs as markers of stance. The author explores variation in the use of reporting verbs in relation to several factors: the writers' L1 (native and non-native speakers), gender (male and female), and degree of professional erudition and experience (professionals and novices). She also discusses the impact of a particular academic culture on writing.

Lindgrén compares BA and MA theses in two fields, linguistics and literature, in terms of academic words, using two prominent academic word lists, the AWL (Coxhead, 2000) and the NAWL (Browne et al., 2013). She correlates lexical parameters with the readability levels of the theses in question, measured by word length and sentence length. The author notes major differences between the two fields of study, and concludes that authors of linguistic final projects may benefit more from the academic word lists. Surprisingly, BA linguistic theses were found to be more difficult to read than their MA counterparts, which may follow from somewhat inappropriate (over)use of certain features.

Part III, entitled Collocations and Lexical Bundles, consists of two studies which examine syntagmatic relationships in lexis, an area which is frequently deemed to be demanding even for high-level proficiency L2 learners.

Erman compares two studies exploring different lexical features in argumentative essays by native and non-native undergraduates at the BA and MA level. The chapter compares the results of manual extraction of collocations and a computer-driven method for retrieval of four-word lexical bundles. The quantitative results are matched with qualitative analyses. In dealing with lexical bundles, the study follows the functional classification devised by Biber et al. (2004), but the author points out the multifunctionality of numerous bundles. The collocation study proposes and applies categories of collocations which refine the picture of the informants' lexical knowledge. Interestingly, the results of the two studies converge in some respects and inform each other.

Lehmann examines a corpus of portfolios produced by graduating part-time students, i.e., in-service teacher retrainees, in terms of the types of lexical bundles, delimited as computer-derived frequency-based four-word clusters. Following Biber et al. (2004), a distinction is also made between referential, discourse and stance bundles. The chapter is unique not only in that it deals with the genre of portfolios, but also in that it analyses the writing of a relatively unusual group of subjects, who decided

to upgrade their qualification long after the completion of their previous university studies. The portfolios revealed the distribution of lexical bundles across functions to be different from native speaker patterns, and closer to the spoken academic register.

Together, the chapters in the book offer the reader a variety of routes to a better understanding of non-native lexis. Each of the chapters deals with a distinct topic and is valuable in its own right. We hope that the juxtaposition and interaction of approaches may unlock a discursive space for negotiating meanings and open up stimulating debates, for example on the benefits of cross-disciplinary approaches or on the possibility of exploring various external factors impacting the use of lexis in L2 discourse.

Whatever the particular objectives of individual chapters, and whatever the circumstances in which the studies were carried out, the research projects reported in this volume seem to have uncovered numerous intriguing tendencies and patternings in the lexis of non-native writing, which support or complement each other in various respects. For example, several studies examine, in their own individual ways, the differences between human raters and computerized assessments, i.e., manual and computer-assisted processing of data. Some scrutinize types of lexical errors, or apply parallel measurements, such as various word-lists. Several studies in the volume also uncovered diverse delicate tendencies of non-native writers to misuse, overuse or underuse particular lexical traits, and pointed out the impact of their play-it-safe strategies.

In addition to advancing the existing knowledge of various issues within the broad domain of L2 lexis, the chapters should provide readers with an opportunity to observe how the authors put into action a repertoire of various modern ways, tools and methodologies in order to study the lexical competence and performance of L2 users. They may thereby indirectly also be putting these approaches and methods to test. In response, the individual research instruments, or possibly the whole research toolkit as such, could be sharpened, enriched or refined. In this sense, the studies included in the volume may have a more general value, showing the benefits, limitations and disadvantages of a number of tools, methods, research designs and procedures.

We hope to be able to appeal to a wide variety of readers, including scholars, researchers, specialists, PhD students, foreign language teachers and undergraduates, who share our interest in non-native lexical resources as they are reflected in written discourse. We are grateful to numerous colleagues, advisors, reviewers and proofreaders, without whom the volume would undoubtedly be less interesting. We are particularly grateful

to Dr. Ellen Valle for checking the language of the manuscript. Any remaining slips or errors are, of course, our own responsibility. For all the genuine endeavour to share with readers these state-of-the-art findings, we know that much has yet to be investigated. From this standpoint, the volume may be more appropriately perceived as a testimony to academic research in progress. In any case, we hope our readers will enjoy the book and perhaps find inspiration in it for their own studies in this fascinating area.

Turku, Szeged and Prague, June 2015

Päivi Pietilä, *Katalin Doró* & *Renata Pípalová*

References

Anthony, L. (n.d.). *AntConc.* Computer software. Available from http://www.laurenceanthony.net/software/antconc/.

Biber, D., Conrad, S., & Cortes, V. (2004). 'If you look at…': Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, *36*(1), 1–22.

Browne, C., Culligan, B., & Phillips, J. (2013). *The New Academic Word List*. Version 1.0. http://www.newacademicwordlist.org/.

Cobb, T. (n.d.). *The Compleat Lexical Tutor*. Computer software. Available online at www.lextutor.ca

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

Fan, M. (2008). An exploratory study of collocational use by ESL students – A task based approach. *System*, *37*, 110–123.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647–672.

Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

—. (n.d.). *RANGE*. Computer program. Available: http://www.victoria.ac.nz/lals/about/staff/paul-nation.

NAWL. Available at http://www.newacademicwordlist.org/.

Sinclair, J. (2005). Meaning in the framework of corpus linguistics. In W. Teubert (Ed.), *Lexicographica* (pp. 20-32). Tubingen: Max Niemeyer.

CHAPTER ONE

RESEARCHING VOCABULARY IN L2 WRITING: METHODOLOGICAL ISSUES AND PEDAGOGICAL IMPLICATIONS

KATALIN DORÓ AND PÄIVI PIETILÄ

This review chapter highlights prominent trends in SLA and writing pedagogy in the large body of recent literature on L2 writing and vocabulary, with a focus on the development of research methods and text-analysis tools. The discussion is also extended to research implications and to applications in L2 teaching and writing instruction, based on large corpora versus small classroom-focused research.

Introduction

Expressing one's idea in writing is a complex and challenging task, especially for the non-native writer. Text composition requires the writer to attend simultaneously to thesis statements and the organization of the points to be included, while keeping in mind the audience and their potential reactions to the text. Writers also need to plan, monitor and review their writing constantly. The additional factor that second or foreign language (L2) writers have to keep in mind, more closely than native writers, is the selection of appropriate lexical and syntactic structures, which may distract their focus from their general writing goals. L2 writing requires both writing skills and language proficiency (Weigle, 2013). Depending on which of these two are the focus of attention, we can distinguish between two broad conceptual dimensions of L2 writing: the dimension of learning to write (LW) and that of writing to learn (WL). The latter refers to the practice of using writing to support learning in other areas, such as content classes (Manchón, 2011). Writing assessment,

especially in a school context, may have three different purposes, as Weigle (2013) puts it:

> There are three somewhat different purposes for writing tests, each asking a somewhat different, though related, question about writing performance: (1) Assessing writing (AW)—does the student have skills in text production and revision, knowledge of genre conventions, and an understanding of how to address readers' expectations in writing? (2) Assessing content through writing (ACW)—does the student understand (and display knowledge in writing about) specific content? (3) Assessing language through writing (ALW)—Has the student mastered the second language skills necessary for achieving their rhetorical goals in English? (p. 89).

In accordance with the assessment of language through writing, L2 learners are often given writing tasks as part of proficiency tests and entrance exams to various study programs. Written essays and academic papers are also frequent assignments in higher education, and often constitute degree requirements, which may reflect all three of the above purposes. While the first two purposes of writing apply to native-speaking students, the third is frequently applied with L2 learners. Given the importance and also the challenges of L2 text production, essays and papers written by learners, as well as academic writing produced by scholars, have been investigated for their linguistic features. Among the different aspects of writing evaluation, vocabulary is believed to be one of the strongest measures of text quality.

Lexical issues in second language writing have received growing attention over the last three decades. The investigation of this broad topic lies at the crossroads of SLA, language teaching, discourse analysis, corpus linguistics and writing pedagogy. Researchers have looked at various aspects, including, but not limited to, the following four areas: a) the lexical content of texts by learners or academics in small and large corpora, developing and applying text-analysis tools; b) the writing process itself, including writing strategies, lexical choices, drafting and editing; c) the attitudes and beliefs of writers, language teachers, readers and text raters; and d) the lexical content of teaching L2 writing. While the focus of attention is often similar, the proposed research questions, terminology, methods and conclusions are often very different, even contradictory.

In this introductory chapter, we review some of the most relevant findings and concepts of vocabulary and L2 texts, with a focus on corpora, general lists, computer-based text analysis and annotated essay scoring.

We discuss the main methods of investigation used in empirical research published in English over the past fifteen years, some of the ongoing debates concerning computer-assisted language teaching and scoring, and the pedagogical implications of using lexical computer tools and research results in language teaching and academic writing for L2 learners.

Computer-based text analysis, corpora and general lists

One major line of tradition in analyzing the vocabulary content of L2 text is frequency. This implies that the quality of a text is influenced by the type of words it contains: the use of less frequent words indicates greater writer proficiency, higher general language proficiency and better text quality (e.g., Laufer, 1998; Nation, 2001). However, as Jiménez Catalán and Fitzpatrick (2014) rightly point out, vocabulary choice in L2 may be based on other principles than frequency both in teaching and in language use. These can include reliance on familiar vocabulary (especially in order to avoid possible lexical errors), classroom language, the lexical content of learning materials, or the use of collocations and other multiword units made up of highly frequent words. Nevertheless, automated analysis tools have operated with frequency lists compiled on the basis of various corpora.

Vocabulary analysers

One of the first and most widely used such tools is Nation's RANGE, constructed on the basis of the General Service List (GSL) compiled by West (1953) decades earlier, before the computer era. RANGE breaks down texts into four frequency lists: K1 (the 1000 most frequent words in English), K2 (the next 1000 most frequent words), academic vocabulary, and the remaining off-list words. The adaptation of RANGE for online use, called VocabProfile (VP), to account for the variety of results that the base corpora can produce, offers analyses using several more recently compiled lists and corpora. The experimental versions of the VP use the British National Corpus (BNC), the Corpus of Contemporary American English (COCA) and the Billuroğlu–Neufeld List (BNL), or a combination of these. The latest versions of the VP provide more detailed and specific profiles (at up to 25 frequency bands) of learners' texts, and offer the possibility of comparing parallel results. The VocabProfile calculates percentages of frequency bands, indicates types, tokens, type/token ratios and lexical density figures. It is part of a larger collection of tools, the Compleat Lexical Tutor (Cobb, n.d.); among its other resources, the CLT

offers the possibility of text comparison (shared and different words in two texts), an N-gram analyser, and concordancing.

Apart from RANGE and the Compleat Lexical Tutor, other freely or commercially available text tools exist that have been used in recently published L2 writing research (e.g., WordSmith Tools, Scott, 2012; AntConc, Anthony, n.d.; and lexical density measures, e.g., D_tools and V_Words, collected under the name of_lognostics, Meara, n.d.). Most of these tools work with raw texts that do not require annotation. However, while non-annotated texts cannot distinguish between homonyms and homographs, corpora built from annotated texts offer more fine-grained and richer analyses. The annotation types most often performed for lexical analysis are part-of-speech (POS) tagging and lemmatization. Some tools do the annotation themselves, while others require prior annotation. The Lexical Complexity Analyzer (Lu, 2012), for example, requires previous POS-tagging and lemmatization (for a recent overview see Lu, 2014).

Service lists

The results of frequency-based analysis are highly dependent on the base corpora with which they are operating. Also, in order to ascertain which words in a language are the most frequent, selection criteria need to be careful and based on up-to-date language use. The General Service List (on which RANGE and the first versions of the VocabProfile were based), sixty years after its compilation, has inspired researchers to design new lists. One of the main reasons for doing so is that most of the content of the GSL is out of date; its aim of serving as a core list, and the methodological rigour of its compilation, nevertheless remain an important example to follow. The original list has served both pedagogical and research purposes, such as the compilation of other lists (e.g., the Academic Word List, Coxhead, 2000, or the Academic Collocation List, Ackermann & Chen, 2013), word selection in teaching and textbook writing, and the design of lexical analysis tools. Two groups of researchers, working independently of each other, have recently published new service lists; the New General Service List (NGSL) Version 1.0 by Browne (2013), and the New General Service List (new-GSL) by Brezina and Gablasova (2015). The two teams used different corpora and selection criteria. Since then, an even newer version (1.01) of the NGSL, with a focus on second language learners, has been designed by Browne and his colleagues; it is based on a careful selection of sub-corpora of the Cambridge English Corpus (BEC) and the Cambridge Learner Corpus (CLC), to ensure the generalisability of the list across genres and users

(Browne, 2014). Researchers, including Bogaards (2008) and Brezina and Gablasova (2015), have argued that the selection of lists based on different corpora may yield very different results. The NGSL 1.01 contains 2,801 words, while the new-GSL consists of 2,494 lemmas. As the NGSL is under constant revision and the various improved lists contain different numbers of words, we may wonder whether a final core list will ever be compiled and generally accepted. Further study is needed to compare the degree of overlap between the new service lists.

Learner corpora

Apart from large corpora containing written and/or oral texts mainly from native speakers of a language or a variety of a language, non-native texts have also been compiled for various reasons. Learner corpora, i.e., large electronic collections of L2 texts, can serve the purposes of both research and teaching. According to Granger (2003) they differ from generally used learner text compilations in two main ways: their electronic and systematic set-up, which allows for software analysis, and their size, which is much larger than the texts typically analysed in research in the fields of second language acquisition (SLA) or foreign language teaching (FLT):

> Size is obviously a relative notion. A corpus of 200,000 words is big in the SLA field where researchers usually rely on much smaller samples but minute in the corpus linguistics field at large where recourse to mega-corpora of several hundred million words has become the norm rather than the exception (Granger 2003, p. 465).

Learner corpora also allow for systematic contrastive analysis between native and non-native texts or L2 texts produced by writers with different language backgrounds. Among the various learner corpora, one which needs to be highlighted is the International Corpus of Learner English (ICLE, Granger, Dagneaux, Meunier, & Paquot, 2009), a large selection of essays written by students with various L1s. Using this corpus, Granger and her colleagues have documented systematic variation across L2 texts according to the writers' L1s, and have also pointed to L1 based writing instruction (e.g., Granger, 1998; Granger, Hung, & Petch-Tyson, 2002; Granger & Paquot, 2009). In order to achieve a parallel L1 corpus of argumentative essays of similar length and on similar topics, Granger and her team compiled the Louvain Corpus of Native English Essays (LOCNESS). Another early large learner corpus, initiated by Milton and his colleagues in the 1990s, is the Hong Kong learner corpus. Since then a

number of other larger and well-known learner corpora have been compiled, such as the International Corpus of Learner English (ICLE) already referred to, and the Cambridge Learner Corpus (CLC). Another Cambridge-based learner corpus is the recently compiled EF-Cambridge Open Language Database (EFCAMDAT), a very large collection of learner writings at various proficiency levels and from a great number of different L1 backgrounds. In contrast to most earlier corpora, EFCAMDAT permits a scrutiny of competence development, as it contains written texts from the same individuals from various points in time (see Geertzen, Alexopoulou, & Korhonen, 2013; Alexopoulou, Geertzen, Korhonen, & Meurers, 2015).

Other corpora focus on specific L1 student populations, such as the Japanese English as a Foreign Language Learner (JEFLL) corpus. In some cases even more restricted selection criteria have been applied, such as a specific task or course assignment, resulting in smaller text compilations at single universities; these include the Active Learning of English for Science Students (ALESS) corpus at the University of Tokyo (Allen, 2009), the Janus Pannonius University (JPU) Corpus in Hungary (Horváth, 2001), or the BATMAT Corpus at Åbo Akademi University in Turku, Finland (Lindgrén, Chapter 7 in this book). These learner corpora of academic English are either used on their own in the investigation of certain linguistic and discourse features (such as lexical bundles, verb use, connectors or pronoun selection) or in comparison with the British Academic Written English (BAWE) corpus. For a detailed description of the BAWE, see Nesi and Gardner (2012) and Gardner and Nesi (2013); a list of learner corpora around the world, is available at the Centre for English Corpus Linguistics in Louvain-la-Neuve, Belgium.

While early learner corpora focused more on general student texts such as argumentative essays, more recent corpora have compiled various academic writing assignments by students in higher education. Two large recent European projects are the Varieties of English for Specific Purposes dAtabase (VESPA) at Louvain-la-Neuve (Granger & Paquot, 2013) and the Corpus of Academic Learner English (CALE) at the University of Bremen (Callies & Zaytseva, 2013; Flowerdew, 2014). Learner corpora have facilitated interlanguage studies, and in recent years have produced a large body of local studies for teaching purposes as well as published research (for a review see e.g., Flowerdew, 2014; Granger, Gilquin, & Meunier, 2015).

Learner language, including its lexicon, varies greatly from L1 language use. The quantitative and qualitative differences include word selection, frequency of words, and the frequent misuse, underuse and

overuse of certain words or phrases (e.g., Granger, 2003; Allen, 2009). Frequency-based differences are detectable when the corpora are run through analytical tools such as the above-mentioned WordSmith Tools. As Granger points out, such tools are not very effective at detecting L2-specific errors; she therefore proposes the use of error-tagged learner corpora (Granger, 2003) for research and in computer assisted language learning (CALL) contexts. Error tagging nevertheless detects only misuse, not under- or over-use; these are relative concepts, and are usually compared to either native corpora or other learner corpora.

Pedagogical implications of using vocabulary analysis tools

L2 lexical research has benefited greatly from corpus linguistics tools, most of which are freely available and offer a built-in interface that makes their application very user-friendly. While the value of researchable large corpora is huge, research often focuses on locally produced learner or scholarly texts, both for research purposes and to improve academic writing instruction. Learners benefit from looking not only at samples produced elsewhere, whether by native speakers or by other, often unidentified L2 learners, but also at texts written by their peers with similar language backgrounds. Smaller-scale studies often have the ultimate goal of turning research outcomes into teaching; yet this aim is often unfulfilled or goes undetected. Researchers who also act as writing instructors, or who have direct contact with students (e.g., through content class teaching, general language development or the rating of student essays and academic texts), can more easily turn results into teaching practice and assessment.

Using corpora and corpus tools for immediate pedagogical purposes (selecting teaching materials, identifying recurring patterns of learner errors, assessing student texts for vocabulary issues, asking students to use tools themselves to check their progress) has great potential. With the availability of user-friendly online tools that are easy to access and operate, and with most L2 texts nowadays being produced in electronic format, this is an attainable goal in academic or essay writing pedagogy. Some initial training in the use of corpus tools is naturally needed for both teachers and students before they can conduct their own searches. Concordancers and frequency analysers can be used to prepare exercises (such as gap-fills), using texts chosen by the instructors. With a data-driven learning approach in mind, the Tex-Lex comparison tool for example of the Compleat Lexical Tutor can be used by students to compare their own texts on similar topics, for example at the beginning

and end of a course or study program to check their lexical progress. Concordancers used in advanced level writing classes, with the assistance of the instructor, may draw students' attention to recurrent patterns of use and errors. They can also study the same selection of lexical items in a closely matching native corpus to see the context in which the given words occur.

Years ago, Berry (1994) and Allan (2002) pointed out another important reason for teachers to use corpus research, namely to develop their own English proficiency and their awareness of language use and vocabulary patterns typical of certain genres and text types. The authors sum up these benefits: frequency lists and concordance lines give teachers better intuitions and develop their analytical skills. Research tools also promote teacher-initiated action research and boost instructors' motivation to enhance their own learning through constant exploration of corpus texts.

Automated essay scoring vs. human raters

The text-analysis tools discussed in the previous section mainly serve research purposes, although they can be adopted in language teaching with advanced level students. With a more practical goal in mind, as a large number of written essays are nowadays produced in proficiency tests and entrance examinations, a need has emerged for automated essay scoring.

Automated essay scoring: methods and validity debates

Kyle and Crossley (2014), in their overview of automated essay scoring (AES), point to its advantages: efficiency of time and cost, reliability, and applicability in both classroom and large-scale testing. However, they also review studies that voice concerns over the limited argumentative and expository essay genres that the AES is able to successfully handle, and its failure to take into account such aspects as argumentation, function, audience, and rhetorical devices. Even with the emergence of automated scoring, the need persists for human raters. The AES systems reviewed by the authors (e-rater(R), Burstein, 2003; IntelliMetric, Rudner, Garcia, & Welch, 2006; Intelligent Essay Assessor, Landauer, Laham, & Foltz, 2003; Writing Pal, McNamara, Crossley, & Roscoe, 2013) all rely on an initial human rating process of sample texts to create a scoring model for the essay prompts. In order to then design statistical models, the essays are further analysed by the AES systems not only for lexical sophistication, grammatical accuracy, and syntactic complexity, but also for rhetorical features and cohesion (Kyle &

Crossley, 2014). Attali and Burstein (2006) evaluated AES systems as opposed to human rating as follows:

> AES systems do not actually read and understand essays as humans do. Whereas human raters may directly evaluate various intrinsic variables of interest, such as diction, fluency, and grammar, in order to produce an essay score, AES systems use approximations or possible correlates of these intrinsic variables. (p. 3)

There is some question as to how accurately these systems can measure the complexity of features that a human rater takes into account while reading. Some studies have therefore compared the results of human raters and AES systems. High correlation figures (ranging between r = .7 and .85) have been reported (for an overview see Kyle & Crossley, 2014). It has also been concluded that AES results are strong predictors of learner proficiency if used together with other measures, such as oral exam points and general study grades (Rudner et al., 2006). In terms of the lexicon of texts, AES systems are able to provide scores for word choice, lexical range, idiomaticity and spelling. They can serve as a general measure in large-scale testing situations, such as the TOEFL proficiency exam, or as a quick first check by instructors or students in a classroom setting, but they cannot and should not replace the human rater if essay scoring forms part of writing development.

The growing interest in and application of automated scoring over the last fifteen years has generated vigorous debate and has led to a large body of published research. There have been special issues of journals, such as the automated assessment writing issue in 2013 of the journal *Assessing Writing*, and numerous articles in writing and assessment journals, along with books and edited volumes (see e.g. Cotos, 2014). One of the major issues with regard to AES is its validity in scoring and predicting writer performance. The focus in the above-mentioned special issue of *Assessing Writing* was on this main topic. In one of the articles, Weigle (2013) summarizes the validity argument with regard to five specific aspects: evaluation, generalization, explanation, extrapolation, and utility. The *evaluation* argument assesses how accurately the scores represent a given performance, usually based on a comparison of human and computer-calculated scores. *Generalization* in the validity debate refers to the degree to which the score assigned by human raters or calculated by the computer would be the same across tasks, raters and occasions of writing. *Explanation* refers to the ability to attribute the score to given constructs. This is often difficult in both AES and human rating: for example a misused preposition may be taken as either a lexical or a syntactic error.

Extrapolation refers to the degree to which scores obtained for a writing task indicate general writing ability or language proficiency. This may well depend on the purpose of the writing test/task, such as an assignment in composition class, a course in academic writing, or writing for professional purposes. Finally, *utility* refers to the usefulness of the scores for students and in writing pedagogy. It may also indicate the long-term impact that scores have on decision making, such as syllabus design, exam scoring and evaluation, or the design of materials for testing and teaching.

Pedagogical implementation of automated essay evaluation in ESL/EFL writing instruction

AES has been used to complement teachers' feedback on writing tasks in both native and non-native contexts. As opposed to the general validity debate, recent empirical studies on the role of automated writing evaluation in ESL/EFL classrooms are very few in number. Li, Link, Ma, Yang and Hegelheimer (2014) carried out a longitudinal mixed-method analysis in three university ESL writing courses in the United States. Although the authors looked at holistic scores and not lexical measures in particular, the implications of the study for classroom practice are valuable and relevant to the present topic. The study looked both at the use of an AES tool during the course and at its interpretation and evaluation by students and teachers. The authors point out that instructors used the tool in three ways: for forewarning, for benchmarking, and as an assessment tool. The first refers to the identification of students with low scores and an alert to weak students that they needed to improve. The second use of AES was to set a threshold score before submission of a paper: low scores were viewed as low quality of writing. The third was the assessment of exams and term papers, although some concern was voiced by the instructors in terms of reliability. AES was found to be motivating for some students, while others were aware of the limitations of the scores and were even doubtful of their interpretation. The study highlighted the pedagogical value of AES: it pushes students to be more independent learners and to improve their writing. Similar writing-classroom implementation, the motivational role and some scepticism towards AES in ESL/EFL contexts has also been documented by Chen and Cheng (2008) in Taiwan and by Link, Dursun, Karakaya and Hegelheimer (2014) and Li, Link, Ma, Yang and Hegelheimer (2014) in the US. In another study, focusing on teachers' AES practices in writing classes, on students' view of its role, and on the effect of AES on the development of writing accuracy between the first and final drafts, Li, Link and Hegelheimer

(2015) also found that AES is mostly viewed favourably by both students and instructors. The authors also point out that the way instructors view AES has a direct impact on how students treat this method of getting feedback. The study calls for further research concerning the full potential of the incorporation and long-term use of automated writing evaluation in writing classes. Notwithstanding the criticism and concerns surrounding AES, the authors are hopeful of its development and integration in course work. However, they also point out the need for human rating and direct feedback from instructors.

Essay evaluation by human raters

Automated text scoring is useful in large testing situations and as an indicator of text quality, for learners' use at home or by writing instructors as part of the course work. Nevertheless, human rating remains a practice in most small-scale situations, such as smaller proficiency exams and class grading. It is a crucially important issue how people read, understand and score essays, how consistent they are in their rating, and how valid and reliable their rating is.

Crossley and his colleagues rightly point out that while selected linguistic features in L2 texts (such as lexical sophistication, grammatical accuracy, syntactic complexity, lexical errors, and the accuracy of syntactic structures) are used for teaching and proficiency measurement, there is little agreement as to how these features influence human raters when they read texts (Crossley, Kyle, Allen, Guo, & McNamara, 2014). It is quite difficult to draw conclusions based on the findings of single studies, or to compare writers and their texts in these studies. Based on the results of earlier research on the factors differentiating texts by more proficient L2 writers from those written by less proficient ones, Crossley and McNamara (2012) hypothesized that writers who had been judged by human raters as highly proficient would manifest more cohesive devices and higher linguistic sophistication in their essays. Cohesion and linguistic sophistication were measured with Coh-Metrix (a computational tool specifically designed to score various aspects of cohesion and linguistic sophistication in written samples). Contrary to expectation, their results indicated that those writers who had been judged as more proficient used fewer cohesive devices than less proficient ones. Various lexical measures (lexical diversity, word frequency, word meaningfulness, and word familiarity), on the other hand, proved highly significant characteristics of better writing quality (p. 131). It is evident that more research is needed to explore what factors actually influence human judgments.

Concluding remarks

Over the last few decades, research into vocabulary issues in learner language, including L2 writing, has taken huge steps forward. This is largely due to the development of new research tools and methods. As long as texts are prepared appropriately (pruned according to the instructions given or to the researcher's plans), their submission to the software is fast and easy, and the results can be obtained in a matter of minutes. Moreover, in addition to software for lexical analysis, there are also various corpora available for researchers to use. Unless the focus is on locally produced texts, it is no longer necessary to compile one's own corpus to study L2 writing tendencies, even though that too remains a viable option.

In this chapter, we provided an overview of certain current tendencies in research into lexical aspects of second language writing, introducing some of the available software, corpora, and word lists. In addition to the research perspectives of these key issues, we also discussed some pedagogical applications and implications for the teaching of writing skills. Based on the recent publications we have reviewed, it is evident that learner corpora, text tools and automated essay evaluation can be successfully incorporated into writing pedagogy. This, however, requires the availability of computer-assisted language learning, a willingness on the part of instructors to experiment with new methods of feedback and evaluation, and some practice using the tools and understanding their output data. We concluded this chapter by considering the role of human raters vis-à-vis automated essay scoring. It seems safe to say that both have their advantages, but should be carefully chosen for specific teaching, research or assessment purposes.

References

Ackermann, K., & Chen, Y. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*, 235–247.

Alexopoulou, T., Geertzen, J., Korhonen, A., & Meurers, D. (2015). Exploring big educational learner corpora for SLA research; perspectives on relative clauses. *International Journal of Learner Corpus Research, 1*(1), 96–129.

Allan, Q. G. (2002). The TELEC secondary learner corpus. A resource for teacher development. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign*

language teaching (pp. 195–211). Philadelphia: John Benjamins Publishing.

Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education, 10*(1), 105–127.

Anthony, L. (n.d.). *AntConc*. Computer software. Available from http://www.laurenceanthony.net/software.html.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® v. 2. *The Journal of Technology, Learning and Assessment 4*(3), 3–30.

Berry, R. (1994). Using concordance printouts for language awareness training. In C. S. Li, D. Mahoney, & J. Richards (Eds.), *Exploring second language teacher development* (pp. 195–208). Hong Kong: City University Press.

Bogaards, P. (2008). Frequency in learners' dictionaries. In E. Bernal, & J. DeCesaris (Eds.), *Proceedings of the XIII EURALEX International Congress, Barcelona* (pp. 1231–1236). Barcelona: UILA, Documenta Universitaria.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, *36*(1), 1–22.

Browne, C. (2013). The New General Service List: Celebrating 60 years of vocabulary learning. *The Language Teacher, 7*(34), 13–16.

—. (2014). The New General Service List Version 1.01: Getting better all the time. *Korea TESOL Journal 11*(1), 35–50.

Burstein, J. (2003). The e-rater scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary approach* (pp. 113–121). Mahwah, NJ: Lawrence Erlbaum Associates.

Callies, M., & Zaytseva, E. (2013). The Corpus of Academic Learner English (CALE) – A new resource for the study and assessment of advanced language proficiency. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty years of learner corpus research: Looking back, moving ahead. Corpora and language in use – Proceedings 1* (pp. 49–59). Louvain-la-Neuve: Presses universitaires de Louvain.

Learner corpus bibliography (n.d.) Centre for English Corpus Linguistics, University of Louvain, Belgium http://www.uclouvain.be/en-cecl-lcworld.html.

Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94–112.

Cobb, T. (n.d.). *The Compleat Lexical Tutor*. Computer software. Available online at www.lextutor.ca.

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement.* Houndmills: Palgrave Macmillan.

Coxhead, A. (2000). The new academic word list. *TESOL Quarterly, 34*(2), 213–238.

Crossley, S. A., Kyle, K., Allen, L. K., Guo, L., & McNamara, D. S. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in automated writing evaluation. *Journal of Writing Assessment, 7*(1). Available from http://journalofwritingassessment.org/article.php?article=74.

Flowerdew, L. (2014). Learner corpus research in EAP: Some key issues and future pathways. *English Language and Linguistics, 20*(2), 43–59.

Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics, 34*(1), 1–29.

Geertzen, J., Alexopoulou, T., & Korhonen, A. (2013). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In R.T. Miller, K.I. Martin, C.M. Eddingon, A. Henery, N. Marcos Miguel, A.M. Tseng, A. Tuninetti, & D. Walter (Eds.), *Proceedings of the 31st Second Language Research Forum (SLRF), Carnegie Mellon* (pp. 240–254). Cascadilla Proceedings Project.

Granger, S. (Ed.). (1998). *Learner English on computer*. London: Longman.

Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal, 20*(3), 465–480.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *The international corpus of learner English. Version 2. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S., Gilquin, G., & Meunier, F. (Eds.), (2015). *The Cambridge handbook of learner corpus research.* Cambridge: Cambridge University Press.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds.), (2002). *Computer learner corpora, second language acquisition, and foreign language teaching*. Amsterdam: Benjamins.

Granger, S., & Paquot, M. (2009). Lexical verbs in academic discourse: A Corpus driven study of learner use. In C. Maggie, S. Hunston, & D. Pecorari (Eds.), *Academic writing: At the interface of corpus and discourse continuum* (pp. 193–214). London: International Publishing.